PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 4 : A61K 39/12, C12Q 1/70 C12N 15/00, C07K 7/10 A61K 37/04, 37/02, C07H 15/12

A1

(11) International Publication Number: WO 89/ 04669

(43) International Publication Date: 1 June 1989 (01.06.89)

(21) International Application Number: PCT/US88/04125

(22) International Filing Date: 18 November 1988 (18.11.88)

(31) Priority Application Numbers: 122,714; 139,886; 161,072; 191,263; Not furnished; Not furnished

(32) Priority Dates: 18 November 1987 (18.11.87); 30 December 1987 (30.12.87); 26 February 1988 (26.02.88); 6 May 1988 (06.05.88); 26 October 1988 (26.10.88); 14 November 1988 (14.11.88)

(33) Priority Country: US

(60) Parent Application or Grant
(63) Related by Continuation
US Not furnished (CIP)
Filed on 14 November 1988 (14.11.88)

(71) Applicant *(for all designated States except US)*: CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : HOUGHTON, Michael [GB/US]; 53 Rosemead Court, Danville, CA 94526 (US). CHOO, Qui-Kim [SG/US]; 5700 Fern Street, El Cerrito, CA 94530 (US). KUO, George [US/US]; 1370 Sixth Avenue, San Francisco, CA 94122 (US).

(74) Agents: MONROY, Gladys, H. et al.; Ciotti & Murashige, Irell & Manella, 545 Middlefield Road, Suite 200, Menlo Park, CA 94025 (US).

(81) Designated States: AT, AT (European patent), AU, BB, BE (European patent), BG, BJ (OAPI patent), BR, CF (OAPI patent), CG (OAPI patent), CH, CH (European patent), CM (OAPI patent), DE, DE (European patent), DK, FI, FR (European patent), GA (OAPI patent), GB, GB (European patent), HU, IT (European patent), JP, KP, KR, LK, LU, LU (European patent), MC, MG, ML (OAPI patent), MR (OAPI patent), MW, NL, NL (European patent), NO, RO, SD, SE, SE (European patent), SN (OAPI patent), SU, TD (OAPI patent), TG (OAPI patent), US.

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: NANBV DIAGNOSTICS AND VACCINES

(57) Abstract

A family of cDNA sequences derived from hepatitis C virus (HCV) are provided. These sequences encode antigens which react immunologically with antibodies present in individuals with non-A non-B hepatitis (NANBH), but which generally are absent from individuals infected with hepatitis A virus (HAV) or hepatitis B virus (HBV), and also are absent from control individuals. A comparison of these cDNA sequences with the sequences in Genebank, and with the sequences of hepatitis delta virus (HDV) and HBV shows a lack of substantial homology. A comparison of the sequences of amino acids encoded in the cDNA with the sequences of Flaviviruses indicates that HCV is a Flavivirus or Flavi-like virus. The HCV cDNA sequences are useful for the design of polynucleotide probes, and for the synthesis of polypeptides which may be used in immunoassays. Both the polynucleotide probes and the polypeptides may be useful for the diagnosis of HCV-induced NANBH, and for screening blood bank specimens and donors for HCV infection. In addition, these cDNA sequences may be useful for the synthesis of immunogenic polypeptides which may be used in vaccines for the treatment, prophylactic and/or therapeutic, of HCV infection. Polypeptides encoded within the cDNA sequences may also be used to raise antibodies against HCV antigenes, and for the purification of antibodies directed against HCV antigenes. These antibodies may be useful in immunoassays for detecting HCV antigenes associated with NANBH in individuals, and in blood bank donations. Moreover, these antibodies may be used for treatment of NANBH in individuals. The reagents provided in the invention also enable the isolation of NANBH agent(s), and the propagation of these agent(s) in tissue culture systems. Moreover, they provide reagents which are useful for screening for antiviral agents for HCV, particularly in tissue culture or animal model systems.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	ML	Mali
AU	Australia	GA	Gabon	MR	Mauritania
BB	Barbados	GB	United Kingdom	MW	Malawi
BE	Belgium	HU	Hungary	NL	Netherlands
BG	Bulgaria	IT	Italy	NO	Norway
BJ	Benin	JP	Japan	RO	Romania
BR	Brazil	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	LI	Liechtenstein	SN	Senegal
CH	Switzerland	LK	Sri Lanka	SU	Soviet Union
CM	Cameroon	LU	Luxembourg	TD	Chad
DE	Germany, Federal Republic of	MC	Monaco	TG	Togo
DK	Denmark	MG	Madagascar	US	United States of America
FI	Finland				

NANBV DIAGNOSTICS AND VACCINES

Technical Field

The invention relates to materials and methodologies for managing the spread of non-A, non-B hepatitis virus (NANBV) infection. More specifically, it relates to diagnostic DNA fragments, diagnostic proteins, diagnostic antibodies and protective antigens and antibodies for an etiologic agent of NANB hepatitis, i.e., hepatitis C virus.

References Cited in the Application

Barr et al. (1986), Biotechniques 4:428.
Botstein (1979), Gene 8:17.
Brinton, M.A. (1986) in THE VIRUSES: THE TOGAVIRIDAE AND FLAVIVIRIDAE (Series eds. Fraenkel-Conrat and Wagner, vol.

eds. Schlesinger and Schlesinger, Plenum Press), p.327-374.

Broach (1981) in: Molecular Biology of the Yeast Saccharomyces, Vol. 1, p.445, Cold Spring Harbor Press.

Broach et al. (1983), Meth. Enz. 101:307.

Chang et al. (1977), Nature 198:1056.

Chirgwin et al. (1979), Biochemistry 18:5294.

Chomczynski and Sacchi (1987), Analytical Biochemistry 162:156.

Clewell et al. (1969), Proc. Natl. Acad. Sci. USA 62:1159.

Clewell (1972), J. Bacteriol. 110:667.

Cohen (1972), Proc. Natl. Acad. Sci. USA 69:2110.

Cousens et al. (1987), gene 61:265.

De Boer et al. (1983), Proc. Natl. Acad. Sci. USA 292:128.

Dreesman et al. (1985), J. Infect. Disease 151:761.

Feinstone, S.M. and Hoofnagle, J.H. (1984), New Engl. J. Med. 311:185.

Fields & Knipe (1986), FUNDAMENTAL VIROLOGY (Raven Press, N.Y.).

Fiers et al. (1978), Nature 273:113.

Gerety, R.J. et al., in VIRAL HEPATITIS AND LIVER DISEASE (Vyas, B.N., Dienstag, J.L., and Hoofnagle, J.H., eds, Grune and Stratton, Inc., 1984) pp 23-47.

Goeddel et al. (1980), Nucleic Acids Res. 8:4057.

Graham and Van der Eb (1978), Virology 52:546.

Grunstein and Hogness (1975), Proc. Natl. Acad. Sci. USA 73:3961.

Grych et al. (1985), Nature 316:74.

Gubler and Hoffman (1983), Gene 25:263.

Hammerling et al. (1981), MONOCLONAL ANTIBODIES AND T-CELL HYBRIDOMAS.

Hess et al. (1968), J. Adv. Enzyme Reg 7:149.

Hinnen et al. (1978), Proc. Natl. Acad. Sci. 75:1929.

Hitzeman et al. (1980), J. Biol. Chem. 255:2073.

Holland et al. (1978), Biochemistry 17:4900.

Holland (1981), J. Biol. Chem. 256: 1385.
Houghton et al. (1981), Nucleic Acids Res. 9:247
Hunyh, T.V. et al. (1985) in DNA CLONING TECHNIQUES; A PRACTICAL APPROACH (D. Glover, Ed., IRL Press, Oxford, U.K.) pp. 49-78.
Immun. Rev. (1982) 62:185.
Iwarson (1987), British Medical J. 295:946.
Kennett et al. (1980) MONOCLONAL ANTIBODIES.
Laemmli (1970), Nature 227, 680.
Lee et al. (1988), Science 239:1288.
Maniatis, T., et al. (1982) MOLECULAR CLONING; A LABORATORY MANUAL (Cold Spring Harbor Press, Cold Spring Harbor, N.Y.).
Mayer and Walker, eds. (1987), IMMUNOCHEMICAL METHODS IN CELL AND MOLECULAR BIOLOGY (Academic Press, London).
Maxam et al. (1980), Methods in Enzymology 65:499.
MacNamara et al. (1984), Science 226:1325.
Messing et al. (1981), Nucleic Acids Res. 9:309.
Messing (1983), Methods in Enzymology 101:20-37.
METHODS IN ENZYMOLOGY (Academic Press).
Michelle et al., Int. Symposium on Viral Hepatitis.
Monath (1986) in THE VIRUSES: THE TOGAVIRADAE AND FLAVIVIRIDAE (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press), p.375-440.
Nagahuma et al. (1984), Anal. Biochem. 141:74.
Neurath et al. (1984), Science 224:392.
Nisonoff et al. (1981), Clin. Immunol. Immunopathol. 21:397-406.
Overby, L.R. (1985), Curr. Hepatol. 5:49.
Overby, L.R. (1986), Curr. Hepatol. 6:65.
Overby, L.R. (1987), Curr. Hepatol. 7:35.
Peleg (1969), Nature 221:193.

Pfefferkorn and Shapiro (1974), in COMPREHENSIVE VIROLOGY, Vol. 2 (Fraenkel-Conrat & Wagner, eds., Plenum, N.Y.) pp. 171-230.
Prince, A.M. (1983), Annu. Rev. Microbiol. 37:217.
Rice et al. (1986) in THE VIRUSES: THE TOGAVIRIDAE AND FLAVIVIRIDAE (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press), p.279-328.
Roehrig (1986) in THE VIRUSES: THE TOGAVIRIDAE AND FLAVIVIRIDAE (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press)
Sadler et al. (1980), Gene 8, 279.
Saiki et al. (1986), Nature 324: 163.
Sanger et al. (1977), Proc. Natl. Acad. Sci. USA 74:5463.
Schlesinger et al. (1986), J. Virol. 60:1153.
Schreier, M., et al. (1980) HYBRIDOMA TECHNIQUES
Scopes (1984), PROTEIN PURIFICATION, PRINCIPLES AND PRACTICE, SECOND EDITION (Springer-Verlag, N.Y.).
Shimatake et al. (1981), Nature 292:128.
Steimer et al. (1986), J. Virol. 58:9.
Stollar (1980), in THE TOGAVIRUSES (R.W. Schlesinger, ed., Academic Press, N.Y.), pp. 584-622.
Taylor et al. (1976), Biochem. Biophys. Acta 442:324.
Towbin et al. (1979), Proc. Natl. Acad. Sci. USA 76, 4350.
Tsu and Herzenberg (1980), in SELECTED METHODS IN CELLULAR IMMUNOLOGY (W.H. Freeman and Co.) pp. 373-391.
Vytdehaag et al. (1985), J. Immunol. 134:1225.
Valenzuela, P., et al. (1982), Nature 298:344.
Valenzuela, P., et al. (1984), in HEPATITIS B (Millman, I., et al., ed, Plenum Press) pp. 225-236.
Warner (1984), DNA 3:401.
Wu and Grossman (1987), Methods in Enzymology Vol. 154, RECOMBINANT DNA, Part E.

Wu (1987), Methods in Enzymology vol 155, RECOMBINANT DNA, part F.
Zoller (1982), Nucleic Acids Res. 10:6487.

Cited Patents

U.S. Patent No. 4,341,761
U.S. Patent No. 4,399,121
U.S. Patent No. 4,427,783
U.S. Patent No. 4,444,887
U.S. Patent No. 4,466,917
U.S. Patent No. 4,472,500
U.S. Patent No. 4,491,632
U.S. Patent No. 4,493,890

Background Art

Non-A, Non-B hepatitis (NANBH) is a transmissible disease or family of diseases that are believed to be viral-induced, and that are distinguishable from other forms of viral-associated liver diseases, including that caused by the known hepatitis viruses, i.e., hepatitis A virus (HAV), hepatitis B virus (HBV), and delta hepatitis virus (HDV), as well as the hepatitis induced by cytomegalovirus (CMV) or Epstein-Barr virus (EBV). NANBH was first identified in transfused individuals. Transmission from man to chimpanzee and serial passage in chimpanzees provided evidence that NANBH is due to a transmissible infectious agent or agents. However, the transmissible agent responsible for NANBH is still unidentified and the number of agents which are causative of the disease are unknown.

Epidemiologic evidence is suggestive that there may be three types of NANBH: the water-borne epidemic type; the blood or needle associated type; and the sporadically occurring (community acquired) type.

However, the number of agents which may be the causative of NANBH are unknown.

Clinical diagnosis and identification of NANBH has been accomplished primarily by exclusion of other viral markers. Among the methods used to detect putative NANBV antigens and antibodies are agar-gel diffusion, counterimmunoelectrophoresis, immunofluorescence microscopy, immune electron microscopy, radioimmunoassay, and enzyme-linked immunosorbent assay. However, none of these assays has proved to be sufficiently sensitive, specific, and reproducible to be used as a diagnostic test for NANBH.

Until now there has been neither clarity nor agreement as to the identity or specificity of the antigen antibody systems associated with agents of NANBH. This is due, at least in part, to the prior or co-infection of HBV with NANBV in individuals, and to the known complexity of the soluble and particulate antigens associated with HBV, as well as to the integration of HBV DNA into the genome of liver cells. In addition, there is the possibility that NANBH is caused by more than one infectious agent, as well as the possibility that NANBH has been misdiagnosed. Moreover, it is unclear what the serological assays detect in the serum of patients with NANBH. It has been postulated that the agar-gel diffusion and counterimmunoelectrophoresis assays detect autoimmune responses or non-specific protein interactions that sometimes occur between serum specimens, and that they do not represent specific NANBV antigen-antibody reactions. The immunofluorescence, and enzyme-linked immunosorbent, and radioimmunoassays appear to detect low levels of a rheumatoid-factor-like material that is frequently present in the serum of patients with NANBH as well as in patients with other hepatic and nonhepatic diseases. Some of the reactivity

detected may represent antibody to host-determined cytoplasmic antigens.

There are a number of candidate NANBV. See, for example the reviews by Prince (1983), Feinstone and Hoofnagle (1984), and Overby (1985, 1986, 1987) and the article by Iwarson (1987). However, there is no proof that any of these candidates represent the etiological agent of NANBH.

The demand for sensitive, specific methods for screening and identifying carriers of NANBV and NANBV contaminated blood or blood products is significant. Post-transfusion hepatitis (PTH) occurs in approximately 10% of transfused patients, and NANBH accounts for up to 90% of these cases. The major problem in this disease is the frequent progression to chronic liver damage (25-55%).

Patient care as well as the prevention of transmission of NANBH by blood and blood products or by close personal contact require reliable diagnostic and prognostic tools to detect nucleic acids, antigens and antibodies related to NANBV. In addition, there is also a need for effective vaccines and immunotherapeutic therapeutic agents for the prevention and/or treatment of the disease.

Disclosure of the Invention

The invention pertains to the isolation and characterization of a newly discovered etiologic agent of NANBH, hepatitis C virus (HCV). More specifically, the invention provides a family of cDNA replicas of portions of HCV genome. These cDNA replicas were isolated by a technique which included a novel step of screening expression products from cDNA libraries created from a particulate agent in infected tissue with sera from patients with NANBH to detect newly synthesized antigens derived from the genome of the heretofore unisolated and

uncharacterized viral agent, and of selecting clones which produced products which reacted immunologically only with sera from infected individuals as compared to non-infected individuals.

Studies of the nature of the genome of the HCV, utilizing probes derived from the HCV cDNA, as well as sequence information contained within the HCV cDNA, are suggestive that HCV is a Flavivirus or a Flavi-like virus.

Portions of the cDNA sequences derived from HCV are useful as probes to diagnose the presence of virus in samples, and to isolate naturally occurring variants of the virus. These cDNAs also make available polypeptide sequences of HCV antigens encoded within the HCV genome(s) and permits the production of polypeptides which are useful as standards or reagents in diagnostic tests and/or as components of vaccines. Antibodies, both polyclonal and monoclonal, directed against HCV epitopes contained within these polypeptide sequences are also useful for diagnostic tests, as therapeutic agents, for screening of antiviral agents, and for the isolation of the NANBV agent from which these cDNAs derive. In addition, by utilizing probes derived from these cDNAs it is possible to isolate and sequence other portions of the HCV genome, thus giving rise to additional probes and polypeptides which are useful in the diagnosis and/or treatment, both prophylactic and therapeutic, of NANBH.

Accordingly with respect to polynucleotides, some aspects of the invention are: a purified HCV polynucleotide; a recombinant HCV polynucleotide; a recombinant polynucleotide comprising a sequence derived from an HCV genome or from HCV cDNA; a recombinant polynucleotide encoding an epitope of HCV; a recombinant vector containing the any of the above recombinant polynucleotides, and a host cell transformed with any of these vectors.

Other aspects of the invention are: a recombinant expression system comprising an open reading frame (ORF) of DNA derived from an HCV genome or from HCV cDNA, wherein the ORF is operably linked to a control sequence compatible with a desired host, a cell transformed with the recombinant expression system, and a polypeptide produced by the transformed cell.

Still other aspects of the invention are: purified HCV, a preparation of polypeptides from the purified HCV; a purified HCV polypeptide; a purified polypeptide comprising an epitope which is immunologically identifiable with an epitope contained in HCV.

Included aspects of the invention are a recombinant HCV polypeptide; a recombinant polypeptide comprised of a sequence derived from an HCV genome or from HCV cDNA; a recombinant polypeptide comprised of an HCV epitope; and a fusion polypeptide comprised of an HCV polypeptide.

Also included in the invention are a monoclonal antibody directed against an HCV epitope; and a purified preparation of polyclonal antibodies directed against an HCV epitope.

Another aspect of the invention is a particle which is immunogenic against HCV infection comprising a non-HCV polypeptide having an amino acid sequence capable of forming a particle when said sequence is produced in a eukaryotic host, and an HCV epitope.

Still another aspect of the invention is a polynucleotide probe for HCV.

Aspects of the invention which pertain to kits are those for: analyzing samples for the presence of polynucleotides derived from HCV comprising a polynucleotide probe containing a nucleotide sequence from HCV of about 8 or more nucleotides, in a suitable container; analyzing samples for the presence of an HCV

antigen comprising an antibody directed against the HCV antigen to be detected, in a suitable container; analyzing samples for the presence of an antibodies directed against an HCV antigen comprising a polypeptide containing an HCV epitope present in the HCV antigen, in a suitable container.

Other aspects of the invention are: a polypeptide comprised of an HCV epitope, attached to a solid substrate; and an antibody to an HCV epitope, attached to a solid substrate.

Still other aspects of the invention are: a method for producing a polypeptide containing an HCV epitope comprising incubating host cells transformed with an expression vector containing a sequence encoding a polypeptide containing an HCV epitope under conditions which allow expression of said polypeptide; and a polypeptide containing an HCV epitope produced by this method.

The invention also includes a method for detecting HCV nucleic acids in a sample comprising reacting nucleic acids of the sample with a probe for an HCV polynucleotide under conditions which allow the formation of a polynucleotide duplex between the probe and the HCV nucleic acid from the sample; and detecting a polynucleotide duplex which contains the probe.

Immunoassays are also included in the invention. These include an immunoassay for detecting an HCV antigen comprising incubating a sample suspected of containing an HCV antigen with a probe antibody directed against the HCV antigen to be detected under conditions which allow the formation of an antigen-antibody complex; and detecting an antigen-antibody complex containing the probe antibody. An immunoassay for detecting antibodies directed against an HCV antigen comprising incubating a sample suspected of containing anti-HCV antibodies with a probe polypeptide

which contains an epitope of the HCV, under conditions which allow the formation of an antibody-antigen complex; and detecting the antibody-antigen complex containing the probe antigen.

Also included in the invention are vaccines for treatment of HCV infection comprising an immunogenic peptide containing an HCV epitope, or an inactivated preparation of HCV, or an attenuated preparation of HCV.

Another aspect of the invention is a tissue culture grown cell infected with HCV.

Yet another aspect of the invention is a method for producing antibodies to HCV comprising administering to an individual an isolated immunogenic polypeptide containing an HCV epitope in an amount sufficient to produce an immune response.

Still another aspect of the invention is a method for isolating cDNA derived from the genome of an unidentified infectious agent, comprising: (a) providing host cells transformed with expression vectors containing a cDNA library prepared from nucleic acids isolated from tissue infected with the agent and growing said host cells under conditions which allow expression of polypeptide(s) encoded in the cDNA; (b) interacting the expression products of the cDNA with an antibody containing body component of an individual infected with said infectious agent under conditions which allow an immunoreaction, and detecting antibody-antigen complexes formed as a result of the interacting; (c) growing host cells which express polypeptides that form antibody-antigen complexes in step (b) under conditions which allow their growth as individual clones and isolating said clones; (d) growing cells from the clones of (c) under conditions which allow expression of polypeptide(s) encoded within the cDNA, and interacting the expression products with antibody containing body components of individuals other than the

individual in step (a) who are infected with the infectious agent and with control individuals uninfected with the agent, and detecting antibody-antigen complexes formed as a result of the interacting; (e) growing host cells which express polypeptides that form antibody-antigen complexes with antibody containing body components of infected individuals and individuals suspected of being infected, and not with said components of control individuals, under conditions which allow their growth as individual clones and isolating said clones; and (f) isolating the cDNA from the host cell clones of (e).

Brief Description of the Drawings

Fig. 1 shows the double-stranded nucleotide sequence of the HCV cDNA insert in clone 5-1-1, and the putative amino acid sequence of the polypeptide encoded therein.

Fig. 2 shows the homologies of the overlapping HCV cDNA sequences in clones 5-1-1, 81, 1-2, and 91.

Fig. 3 shows a composite sequence of HCV cDNA derived from overlapping clones 81, 1-2, and 91, and the amino acid sequence encoded therein.

Fig. 4 shows the double-stranded nucleotide sequence of the HCV cDNA insert in clone 81, and the putative amino acid sequence of the polypeptide encoded therein.

Fig. 5 shows the HCV cDNA sequence in clone 36, the segment which overlaps the NANBV cDNA of clone 81, and the polypeptide sequence encoded within clone 36.

Fig. 6 shows the combined ORF of HCV cDNAs in clones 36 and 81, and the polypeptide encoded therein.

Fig. 7 shows the HCV cDNA sequence in clone 32, the segment which overlaps clone 81, and the polypeptide encoded therein.

Fig. 8 shows the HCV cDNA sequence in clone 35, the segment which overlaps clone 36, and the polypeptide encoded therein.

Fig. 9 shows the combined ORF of HCV cDNAs in clones 35, 36, 81, and 32, and the polypeptide encoded therein.

Fig. 10 shows the HCV cDNA sequence in clone 37b, the segment which overlaps clone 35, and the polypeptide encoded therein.

Fig. 11 shows the HCV cDNA sequence in clone 33b, the segment which overlaps clone 32, and the polypeptide encoded therein.

Fig. 12 shows the HCV cDNA sequence in clone 40b, the segment which overlaps clone 37b, and the polypeptide encoded therein.

Fig. 13 shows the HCV cDNA sequence in clone 25c, the segment which overlaps clone 33b, and the polypeptide encoded therein.

Fig. 14 shows the nucleotide sequence and polypeptide encoded therein of the ORF which extends through the HCV cDNAs in clones 40b, 37b, 35, 36, 81, 32, 33b, and 25c.

Fig. 15 shows the HCV cDNA sequence in clone 33c, the segment which overlaps clones 40b and 33c, and the amino acids encoded therein.

Fig. 16 shows the HCV cDNA sequence in clone 8h, the segment which overlaps clone 33c, and the amino acids encoded therein.

Fig. 17 shows the HCV cDNA sequence in clone 7e, the segment which overlaps clone 8h, and the amino acids encoded therein.

Fig. 18 shows the HCV cDNA sequence in clone 14c, the segment which overlaps clone 25c, and the amino acids encoded therein.

Fig. 19 shows the HCV cDNA sequence in clone 8f, the segment which overlaps clone 14c, and the amino acids encoded therein.

Fig. 20 shows the HCV cDNA sequence in clone 33f, the segment which overlaps clone 8f, and the amino acids encoded therein.

Fig. 21 shows the HCV cDNA sequence in clone 33g, the segment which overlaps clone 33f, and the amino acids encoded therein.

Fig. 22 shows the HCV cDNA sequence in clone 7f, the segment which overlaps the sequence in clone 7e, and the amino acids encoded therein.

Fig. 23 shows the HCV cDNA sequence in clone 11b, the segment which overlaps the sequence in clone 7f, and the amino acids encoded therein.

Fig. 24 shows the HCV cDNA sequence in clone 14i, the segment which overlaps the sequence in clone 11b, and the amino acids encoded therein.

Fig. 25 shows the HCV cDNA sequence in clone 39c, the segment which overlaps the sequence in clone 33g, and the amino acids encoded therein.

Fig. 26 shows a composite HCV cDNA sequence derived from the aligned cDNAs in clones 14i, 11b, 7f, 7e, 8h, 33c 40b 37b 35 36, 81, 32, 33b, 25c, 14c, 8f, 33f, and 33g; also shown is the amino acid sequence of the polypeptide encoded in the extended ORF in the derived sequence.

Fig. 27 shows the sequence of the HCV cDNA in clone 12f, the segment which overlaps clone 14i, and the amino acids encoded therein.

Fig. 28 shows the sequence of the HCV cDNA in clone 35f, the segment which overlaps clone 39c, and the amino acids encoded therein.

Fig. 29 shows the sequence of the HCV cDNA in clone 19g, the segment which overlaps clone 35f, and the amino acids encoded therein.

Fig. 30 shows the sequence of clone 26g, the segment which overlaps clone 19g, and the amino acids encoded therein.

Fig. 31 shows the sequence of clone 15e, the segment which overlaps clone 26g, and the amino acids encoded therein.

Fig. 32 shows the sequence in a composite cDNA, which was derived by aligning clones 12f through 15e in the 5' to 3' direction; it also shows the amino acids encoded in the continuous ORF.

Fig. 33 shows a photograph of Western blots of a fusion protein, SOD-NANB$_{5-1-1}$, with chimpanzee serum from chimpanzees infected with BB-NANB, HAV, and HBV.

Fig. 34 shows a photograph of Western blots of a fusion protein, SOD-NANB$_{5-1-1}$, with serum from humans infected with NANBV, HAV, HBV, and from control humans..

Fig. 35 is a map showing the significant features of the vector pAB24.

Fig. 36 shows the putative amino acid sequence of the carboxy-terminus of the fusion polypeptide C100-3 and the nucleotide sequence encoding it.

Fig. 37A is a photograph of a coomassie blue stained polyacrylamide gel which identifies C100-3 expressed in yeast.

Fig. 37B shows a Western blot of C100-3 with serum from a NANBV infected human.

Fig. 38 shows an autoradiograph of a Northern blot of RNA isolated from the liver of a BB-NANBV infected chimpanzee, probed with BB-NANBV cDNA of clone 81.

Fig. 39 shows an autoradiograph of NANBV nucleic acid treated with RNase A or DNase I, and probed with BB-NANBV cDNA of clone 81.

Fig. 40 shows an autoradiograph of nucleic acids extracted from NANBV particles captured from infected plasma with anti-NANB$_{5-1-1}$, and probed with ^{32}P-labeled NANBV cDNA from clone 81.

Fig. 41 shows autoradiographs of filters containing isolated NANBV nucleic acids, probed with ^{32}P-labeled plus and minus strand DNA probes derived from NANBV cDNA in clone 81.

Fig. 42 shows the homologies between a polypeptide encoded in HCV cDNA and an NS protein from Dengue flavivirus.

Fig. 43 shows a histogram of the distribution of HCV infection in random samples, as determined by an ELISA screening.

Fig. 44 shows a histogram of the distribution of HCV infection in random samples using two configurations of immunoglobulin-enzyme conjugate in an ELISA assay.

Fig. 45 shows the sequences in a primer mix, derived from a conserved sequence in NS1 of flaviviruses.

Fig. 46 shows the HCV cDNA sequence in clone k9-1, the segment which overlaps the cDNA in Fig. 26, and the amino acids encoded therein.

Fig. 47 shows the sequence in a composite cDNA which was derived by aligning clones k9-1 through 15e in the 5′ to 3′ direction; it also shows the amino acids encoded in the continuous ORF.

Modes for Carrying Out the Invention

I. Definitions

The term "hepatitis C virus" has been reserved by workers in the field for an heretofore unknown etiologic agent of NANBH. Accordingly, as used herein, "hepatitis C virus" (HCV) refers to an agent causitive of NANBH, which was formerly referred to as NANBV and/or BB-

NANBV. The terms HCV, NANBV, and BB-NANBV are used interchangeably herein. As an extension of this terminology, the disease caused by HCV, formerly called NANB hepatitis (NANBH), is called hepatitis C. The terms NANBH and hepatitis C may be used interchangeably herein.

The term "HCV", as used herein, denotes a viral species which causes NANBH, and attenuated strains or defective interfering particles derived therefrom. As shown infra., the HCV genome is comprised of RNA. It is known that RNA containing viruses have relatively high rates of spontaneous mutation, i.e., reportedly on the order of 10^{-3} to 10^{-4} per nucleotide (Fields & Knipe (1986)). Therefore, there are multiple strains within the HCV species described infra. The compositions and methods described herein, enable the propagation, identification, detection, and isolation of the various related strains. Moreover, they also allow the preparation of diagnostics and vaccines for the various strains, and have utility in screening procedures for anti-viral agents for pharmacologic use in that they inhibit replication of HCV.

The information provided herein, although derived from one strain of HCV, hereinafter referred to as CDC/HCV1, is sufficient to allow a viral taxonomist to identify other strains which fall within the species. As described herein, we have discovered that HCV is a Flavivirus or Flavi-like virus. The morphology and composition of Flavivirus particles are known, and are discussed in Brinton (1986). Generally, with respect to morphology, Flaviviruses contain a central nucleocapsid surrounded by a lipid bilayer. Virions are spherical and have a diameter of about 40-50 nm. Their cores are about 25-30 nm in diameter. Along the outer surface of the virion envelope are projections that are about 5-10 nm long with terminal knobs about 2 nm in diameter.

HCV encodes an epitope which is immunologically identifiable with an epitope in the HCV genome from which the cDNAs described herein are derived; preferably the epitope is encoded in a cDNA described herein. The epitope is unique to HCV when compared to other known Flaviviruses. The uniqueness of the epitope may be determined by its immunological reactivity with HCV and lack of immunological reactivity with other Flavivirus species. Methods for determining immunological reactivity are known in the art, for example, by radioimmunoassay, by Elisa assay, by hemagglutination, and several examples of suitable techniques for assays are provided herein.

In addition to the above, the following parameters are applicable, either alone or in combination, in identifying a strain as HCV. Since HCV strains are evolutionarily related, it is expected that the overall homology of the genomes at the nucleotide level will be about 40% or greater, preferably about 60% or greater, and even more preferably about 80% or greater; and in addition that there will be corresponding contiguous sequences of at least about 13 nucleotides. The correspondence between the putative HCV strain genomic sequence and the CDC/CH1 HCV cDNA sequence can be determined by techniques known in the art. For example, they can be determined by a direct comparison of the sequence information of the polynucleotide from the putative HCV, and the HCV cDNA sequence(s) described herein. For example, also, they can be determined by hybridization of the polynucleotides under conditions which form stable duplexes between homologous regions (for example, those which would be used prior to S_1 digestion), followed by digestion with single stranded specific nuclease(s), followed by size determination of the digested fragments.

Because of the evolutionary relationship of the strains of HCV, putative HCV strains are identifiable by

their homology at the polypeptide level. Generally, HCV strains are more than about 40% homologous, preferably more than about 60% homologous, and even more preferably more than about 80% homologous at the polypeptide level. The techniques for determining amino acid sequence homology are known in the art. For example, the amino acid sequence may be determined directly and compared to the sequences provided herein. For example also, the nucleotide sequence of the genomic material of the putative HCV may be determined (usually via a cDNA intermediate); the amino acid sequence encoded therein can be determined, and the corresponding regions compared.

As used herein, a polynucleotide "derived from" a designated sequence, for example, the HCV cDNA, particularly those exemplified in Figs. 1-32, or from an HCV genome, refers to a polynucleotide sequence which is comprised of a sequence of approximately at least about 6 nucleotides, is preferably at least about 8 nucleotides, is more preferably at least about 10-12 nucleotides, and even more preferably at least about 15-20 nucleotides corresponding, i.e., homologous to or complementary to, a region of the designated nucleotide sequence. Preferably, the sequence of the region from which the polynucleotide is derived is homologous to or complementary to a sequence which is unique to an HCV genome. Whether or not a sequence is unique to the HCV genome can be determined by techniques known to those of skill in the art. For example, the sequence can be compared to sequences in databanks, e.g., Genebank, to determine whether it is present in the uninfected host or other organisms. The sequence can also be compared to the known sequences of other viral agents, including those which are known to induce hepatitis, e.g., HAV, HBV, and HDV, and to other members of the Flaviviridae. The correspondence or non-correspondence of the derived sequence to other sequences

can also be determined by hybridization under the appropriate stringency conditions. Hybridization techniques for determining the complementarity of nucleic acid sequences are known in the art, and are discussed infra. See also, for example, Maniatis et al. (1982). In addition, mismatches of duplex polynucleotides formed by hybridization can be determined by known techniques, including for example, digestion with a nuclease such as S1 that specifically digests single-stranded areas in duplex polynucleotides. Regions from which typical DNA sequences may be "derived" include but are not limited to, for example, regions encoding specific epitopes, as well as non-transcribed and/or non-translated regions.

The derived polynucleotide is not necessarily physically derived from the nucleotide sequence shown, but may be generated in any manner, including for example, chemical synthesis or DNA replication or reverse transcription or transcription, which are based on the information provided by the sequence of bases in the region(s) from which the polynucleotide is derived. In addition, combinations of regions corresponding to that of the designated sequence may be modified in ways known in the art to be consistent with an intended use.

Similarly, a polypeptide or amino acid sequence "derived from" a designated nucleic acid sequence, for example, the sequences in Figs. 1-32, or from an HCV genome, refers to a polypeptide having an amino acid sequence identical to that of a polypeptide encoded in the sequence, or a portion thereof wherein the portion consists of at least 3-5 amino acids, and more preferably at least 8-10 amino acids, and even more preferably at least 11-15 amino acids, or which is immunologically identifiable with a polypeptide encoded in the sequence.

A recombinant or derived polypeptide is not necessarily translated from a designated nucleic acid

sequence, for example, the sequences in Figs. 1-26, or from an HCV genome; it may be generated in any manner, including for example, chemical synthesis, or expression of a recombinant expression system, or isolation from mutated HCV.

The term "recombinant polynucleotide" as used herein intends a polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation: (1) is not associated with all or a portion of the polynucleotide with which it is associated in nature or in the form of a library; and/or (2) is linked to a polynucleotide other than that to which it is linked in nature.

The term "polynucleotide" as used herein refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. This term refers only to the primary structure of the molecule. Thus, this term includes double- and single-stranded DNA, as well as double- and single stranded RNA. It also includes modified, for example, by methylation and/or by capping, and unmodified forms of the polynucleotide.

As used herein, the term "HCV containing a sequence corresponding to a cDNA" means that the HCV contains a polynucleotide sequence which is homologous to or complementary to a sequence in the designated DNA; the degree of homology or complementarity to the cDNA will be approximately 50% or greater, will preferably be at least about 70%, and even more preferably will be at least about 90%. The sequences which correspond will be at least about 70 nucleotides, preferably at least about 80 nucleotides, and even more preferably at least about 90 nucleotides in length. The correspondence between the HCV sequence and the cDNA can be determined by techniques known in the art, including, for example, a direct comparison of the sequenced material with the cDNAs

described, or hybridization and digestion with single strand nucleases, followed by size determination of the digested fragments.

The term "purified of viral polynucleotide" refers to an HCV genome or fragment thereof which is essentially free, i.e., contains less than about 50%, preferably less than about 70%, and even more preferably less than about 90% of polypeptides with which the viral polynucleotide is naturally associated. Techniques for purifying viral polynucleotides from viral particles are known in the art, and include for example, disruption of the particle with a chaotropic agent, and separation of the polynucleotide(s) and polypeptides by ion-exchange chromatography, affinity chromatography, and sedimentation according to density.

The term "purified viral polypeptided" refers to an HCV polypeptide or fragment thereof which is essentially free, i.e., contains less than about 50%, preferably less than about 70%, and even more preferably less than about 90%, of cellular components with which the viral polypeptide is naturally associated. Techniques for purifying viral polypeptides are known in the art, and examples of these techniques are discussed infra.

"Recombinant host cells", "host cells", "cells", "cell lines", "cell cultures:, and other such terms denoting microorganisms or higher eukaryotic cell lines cultured as unicellular entities refer to cells which can be, or have been, used as recipients for recombinant vector or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement as the original parent, due to accidental or deliberate mutation. Progeny of the parental cell which are sufficiently similar to the parent

to be characterized by the relevant property, such as the presence of a nucleotide sequence encoding a desired peptide, are included in the progeny intended by this definition, and are covered by the above terms.

A "replicon" is any genetic element, e.g., a plasmid, a chromosome, a virus, that behaves as an autonomous unit of polynucleotide replication within a cell; i.e., capable of replication under its own control.

A "vector" is a replicon in which another polynucleotide segment is attached, so as to bring about the replication and/or expression of the attached segment.

"Control sequence" refers to polynucleotide sequences which are necessary to effect the expression of coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and terminators; in eukaryotes, generally, such control sequences include promoters, terminators and, in some instances, enhancers. The term "control sequences" is intended to include, at a minimum, all components whose presence is necessary for expression, and may also include additional components whose presence is advantageous, for example, leader sequences.

"Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

An "open reading frame" (ORF) is a region of a polynucleotide sequence which encodes a polypeptide; this region may represent a portion of a coding sequence or a total coding sequence.

A "coding sequence" is a polynucleotide sequence which is transcribed into mRNA and/or translated into a polypeptide when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the 5′-terminus and a translation stop codon at the 3′-terminus. A coding sequence can include, but is not limited to mRNA, cDNA, and recombinant polynucleotide sequences.

"Immunologically identifiable with/as" refers to the presence of epitope(s) and polypeptides(s) which are also present in and are unique to the designated polypeptide(s), usually HCV proteins. Immunological identity may be determined by antibody binding and/or competition in binding; these techniques are known to those of average skill in the art, and are also illustrated infra. The uniqueness of an epitope can also be determined by computer searches of known data banks, e.g. Genebank, for the polynucleotide sequences which encode the epitope, and by amino acid sequence comparisons with other known proteins.

As used herein, "epitope" refers to an antigenic determinant of a polypeptide; an epitope could comprise 3 amino acids in a spatial conformation which is unique to the epitope, generally an epitope consists of at least 5 such amino acids, and more usually, consists of at least 8-10 such amino acids. Methods of determining the spatial conformation of amino acids are known in the art, and include, for example, x-ray crystallography and 2-dimensional nuclear magnetic resonance.

A polypeptide is "immunologically reactive" with an antibody when it binds to an antibody due to antibody recognition of a specific epitope contained within the polypeptide. Immunological reactivity may be determined by antibody binding, more particularly by the kinetics of

antibody binding, and/or by competition in binding using as competitor(s) a known polypeptide(s) containing an epitope against which the antibody is directed. The techniques for determining whether a polypeptide is immunologically reactive with an antibody are known in the art.

As used herein, the term "immunogenic polypeptide containing an HCV epitope" includes naturally occurring HCV polypeptides or fragments thereof, as well as polypeptides prepared by other means, for example, chemical synthesis, or the expression of the polypeptide in a recombinant organism.

The term "polypeptide" refers to a molecular chain of amino acids and does not refer to a specific length of the product; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not refer to post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like.

"Transformation", as used herein, refers to the insertion of an exogenous polynucleotide into a host cell, irrespective of the method used for the insertion, for example, direct uptake, transduction, or f-mating. The exogenous polynucleotide may be maintained as a non-integrated vector, for example, a plasmid, or alternatively, may be integrated into the host genome.

"Treatment" as used herein refers to prophylaxis and/or therapy.

An "individual", as used herein, refers to vertebrates, particularly members of the mammalian species, and includes but is not limited to domestic animals, sports animals, primates, and humans.

As used herein, the "plus strand" of a nucleic acid contains the sequence that encodes the polypeptide.

The "minus strand" contains a sequence which is complementary to that of the "plus strand".

As used herein, a "positive stranded genome" of a virus is one in which the genome, whether RNA or DNA, is single-stranded and which encodes a viral polypeptide(s). Examples of positive stranded RNA viruses include Togaviridae, Coronaviridae, Retroviridae, Picornaviridae, and Caliciviridae. Included also, are the Flaviviridae, which were formerly classified as Togaviradae. See Fields & Knipe (1986).

As used herein, "antibody containing body component" refers to a component of an individual's body which is a source of the antibodies of interest. Antibody containing body components are known in the art, and include but are not limited to, for example, plasma, serum, spinal fluid, lymph fluid, the external sections of the respiratory, intestinal, and genitourinary tracts, tears, saliva, milk, white blood cells, and myelomas.

As used herein, "purified HCV" refers to a preparation of HCV which has been isolated from the cellular constituents with which the virus is normally associated, and from other types of viruses which may be present in the infected tissue. The techniques for isolating viruses are known to those of skill in the art, and include, for example, centrifugation and affinity chromatography; a method of preparing purified HCV is discussed infra.

II. Description of the Invention

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Maniatis, Fitsch & Sambrook,

MOLECULAR CLONING; A LABORATORY MANUAL (1982); DNA CLONING, VOLUMES I AND II (D.N Glover ed. 1985); OLIGONUCLEOTIDE SYNTHESIS (M.J. Gait ed, 1984); NUCLEIC ACID HYBRIDIZATION (B.D. Hames & S.J. Higgins eds. 1984); TRANSCRIPTION AND TRANSLATION (B.D. Hames & S.J. Higgins eds. 1984); ANIMAL CELL CULTURE (R.I. Freshney ed. 1986); IMMOBILIZED CELLS AND ENZYMES (IRL Press, 1986); B. Perbal, A PRACTICAL GUIDE TO MOLECULAR CLONING (1984); the series, METHODS IN ENZYMOLOGY (Academic Press, Inc.); GENE TRANSFER VECTORS FOR MAMMALIAN CELLS (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory), Methods in Enzymology Vol. 154 and Vol. 155 (Wu and Grossman, and Wu, eds., respectively), Mayer and Walker, eds. (1987), IMMUNOCHEMICAL METHODS IN CELL AND MOLECULAR BIOLOGY (Academic Press, London), Scopes, (1987), PROTEIN PURIFICATION: PRINCIPLES AND PRACTICE, Second Edition (Springer-Verlag, N.Y.), and HANDBOOK OF EXPERIMENTAL IMMUNOLOGY, VOLUMES I-IV (D.M. Weir and C. C. Blackwell eds 1986).

All patents, patent applications, and publications mentioned herein, both supra and infra, are hereby incorporated herein by reference.

The useful materials and processes of the present invention are made possible by the provision of a family of closely homologous nucleotide sequences isolated from a cDNA library derived from nucleic acid sequences present in the plasma of an HCV infected chimpanzee. This family of nucleotide sequences is not of human or chimpanzee origin, since it hybridizes to neither human nor chimpanzee genomic DNA from uninfected individuals, since nucleotides of this family of sequences are present only in liver and plasma of chimpanzees with HCV infection, and since the sequence is not present in Genebank. In addition, the family of sequences shows no significant homology to sequences contained within the HBV genome.

The sequence of one member of the family, contained within clone 5-1-1, has one continuous open reading frame (ORF) which encodes a polypeptide of approximately 50 amino acids. Sera from HCV infected humans contain antibodies which bind to this polypeptide, whereas sera from non-infected humans do not contain antibodies to this polypeptide. Finally, whereas the sera from uninfected chimpanzees do not contain antibodies to this polypeptide, the antibodies are induced in chimpanzees following acute NANBH infection. Moreover, antibodies to this polypeptide are not detected in chimps and humans infected with HAV and HBV. By these criteria the sequence is a cDNA to a viral sequence, wherein the virus causes or is associated with NANBH; this cDNA sequence is shown in Fig. 1. As discussed infra, the cDNA sequence in clone 5-1-1 differs from that of the other isolated cDNAs in that it contains 28 extra base pairs.

A composite of other identified members of the cDNA family, which were isolated using as a probe a synthetic sequence equivalent to a fragment of the cDNA in clone 5-1-1, is shown in Fig. 3. A member of the cDNA family which was isolated using a synthetic sequence derived from the cDNA in clone 81 is shown in Fig. 5, and the composite of this sequence with that of clone 81 is shown in Fig. 6. Other members of the cDNA family, including those present in clones 12f, 14i, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, and 33g, 39c, 35f, 19g, 26g and 15e are described in Section IV.A.. A composite of the cDNAs in these clones is described in Section IV.A.19, and shown in Fig. 32. The composite cDNA shows that it contains one continuous ORF, and thus encodes a polyprotein. This data is consistent with the suggestion, discussed infra., that HCV is a flavivirus or flavi-like virus.

The availability of this family of cDNAs shown in Figs. 1-32, inclusive, permits the construction of DNA probes and polypeptides useful in diagnosing NANBH due to HCV infection and in screening blood donors as well as donated blood and blood products for infection. For example, from the sequences it is possible to synthesize DNA oligomers of about 8-10 nucleotides, or larger, which are useful as hybridization probes to detect the presence of the viral genome in, for example, sera of subjects suspected of harboring the virus, or for screening donated blood for the presence of the virus. The family of cDNA sequences also allows the design and production of HCV specific polypeptides which are useful as diagnostic reagents for the presence of antibodies raised during NANBH. Antibodies to purified polypeptides derived from the cDNAs may also be used to detect viral antigens in infected individuals and in blood.

Knowledge of these cDNA sequences also enable the design and production of polypeptides which may be used as vaccines against HCV and also for the production of antibodies, which in turn may be used for protection against the disease, and/or for therapy of HCV infected individuals.

Moreover, the family of cDNA sequences enables further characterization of the HCV genome. Polynucleotide probes derived from these sequences may be used to screen cDNA libraries for additional overlapping cDNA sequences, which, in turn, may be used to obtain more overlapping sequences. Unless the genome is segmented and the segments lack common sequences, this technique may be used to gain the sequence of the entire genome. However, if the genome is segmented, other segments of the genome can be obtained by repeating the lambda-gt11 serological screening procedure used to isolate the cDNA clones

described herein, or alternatively by isolating the genome from purified HCV particles.

The family of cDNA sequences and the polypeptides derived from these sequences, as well as antibodies directed against these polypeptides are also useful in the isolation and identification of the BB-NANBV agent(s). For example, antibodies directed against HCV epitopes contained in polypeptides derived from the cDNAs may be used in processes based upon affinity chromatography to isolate the virus. Alternatively, the antibodies may be used to identify viral particles isolated by other techniques. The viral antigens and the genomic material within the isolated viral particles may then be further characterized.

The information obtained from further sequencing of the HCV genome(s), as well as from further characterization of the HCV antigens and characterization of the genome enables the design and synthesis of additional probes and polypeptides and antibodies which may be used for diagnosis, for prevention, and for therapy of HCV induced NANBH, and for screening for infected blood and blood-related products.

The availability of probes for HCV, including antigens and antibodies, and polynucleotides derived from the genome from which the family of cDNAs is derived also allows for the development of tissue culture systems which will be of major use in elucidating the biology of HCV. This in turn, may lead to the development of new treatment regimens based upon antiviral compounds which preferentially inhibit the replication of, or infection by HCV.

The method used to identify and isolate the etiologic agent for NANBH is novel, and it may be applicable to the identification and/or isolation of heretofore uncharacterized agents which contain a genome, and

which are associated with a variety of diseases, including those induced by viruses, viroids, bacteria, fungi and parasites. In this method, a cDNA library was created from the nucleic acids present in infected tissue from an infected individual. The library was created in a vector which allowed the expression of polypeptides encoded in the cDNA. Clones of host cells containing the vector, which expressed an immunologically reactive fragment of a polypeptide of the etiologic agent, were selected by immunological screening of the expression products of the library with an antibody containing body component from another individual previously infected with the putative agent. The steps in the immunological screening technique included interacting the expression products of the cDNA containing vectors with the antibody containing body component of a second infected individual, and detecting the formation of antibody-antigen complexes between the expression product(s) and antibodies of the second infected individual. The isolated clones are screened further immunologically by interacting their expression products with the antibody containing body components of other individuals infected with the putative agent and with control individuals uninfected with the putative agent, and detecting the formation of antigen-antibody complexes with antibodies from the infected individuals; and the cDNA containing vectors which encode polypeptides which react immunologically with antibodies from infected individuals and individuals suspected of being infected with the agent, but not with control individuals are isolated. The infected individuals used for the construction of the cDNA library, and for the immunological screening need not be of the same species.

The cDNAs isolated as a result of this method, and their expression products, and antibodies directed against the expression products, are useful in character-

izing and/or capturing the etiologic agent. As described in more detail infra, this method has been used successfully to isolate a family of cDNAs derived from the HCV genome.

II.A. Preparation of the cDNA Sequence

Pooled serum from a chimpanzee with chronic HCV infection and containing a high titer of the virus, i.e., at least 10^6 chimp infectious doses/ml (CID/ml) was used to isolate viral particles; nucleic acids isolated from these particles was used as the template in the construction of a cDNA library to the viral genome. The procedures for isolation of putative HCV particles and for constructing the cDNA library in lambda-gt11 is discussed in Section IV.A.1. Lambda-gt11 is a vector that has been developed specifically to express inserted cDNAs as fusion polypeptides with beta-galactosidase and to screen large numbers of recombinant phage with specific antisera raised against a defined antigen. The lambda-gt11 cDNA library generated from a cDNA pool containing cDNA of approximate mean size of 200 base pairs was screened for encoded epitopes that could bind specifically with sera derived from patients who had previously experienced NANB hepatitis. Huynh, T.V. et al. (1985). Approximately 10^6 phages were screened, and five positive phages were identified, purified, and then tested for specificity of binding to sera from different humans and chimpanzees previously infected with the HCV agent. One of the phages, 5-1-1, bound 5 of the 8 human sera tested. This binding appeared selective for sera derived from patients with prior NANB hepatitis infections since 7 normal blood donor sera did not exhibit such binding.

The sequence of the cDNA in recombinant phage 5-1-1 was determined, and is shown in Fig. 1. The polypeptide encoded by this cloned cDNA, which is in the

same translational frame as the N-terminal beta-Galactosidase moiety of the fusion polypeptide is shown above the nuclectide sequence. This translational ORF, therefore, encodes an epitope(s) specifically recognized by sera from patients with NANB hepatitis infections.

The availability of the cDNA in recombinant phage 5-1-1 has allowed for the isolation of other clones containing additional segments and/or alternative segments of cDNA to the viral genome. The lambda-gt11 cDNA library described supra, was screened using a synthetic polynucleotide derived from the sequence of the cloned 5-1-1 cDNA. This screening yielded three other clones, which were identified as 81, 1-2 and 91; the cDNAs contained within these clones were sequenced. See Sections IV.A.3. and IV.A.4. The homologies between the four independent clones are shown in Fig. 2, where the homologies are indicated by the vertical lines. Sequences of nucleotides present uniquely in clones 5-1-1, 81, and 91 are indicated by small letters.

The cloned cDNAs present in recombinant phages in clones 5-1-1, 81, 1-2, and 91 are highly homologous, and differ in only two regions. First, nucleotide number 67 in clone 1-2 is a thymidine, whereas the other three clones contain a cytidine residue in this position. This substitution, however, does not alter the nature of the encoded amino acid.

The second difference between the clones is that clone 5-1-1 contains 28 base pairs at its 5'-terminus which are not present in the other clones. The extra sequence may be a 5'-terminal cloning artifact; 5'-terminal cloning artifacts are commonly observed in the products of cDNA methods.

Synthetic sequences derived from the 5'-region and the 3'-region of the HCV cDNA in clone 81 were used to screen and isolate cDNAs from the lambda-gt11 NANBV cDNA

library, which overlapped clone 81 cDNA (Section IV.A.5.). The sequences of the resulting cDNAs, which are in clone 36 and clone 32, respectively, are shown in Fig. 5 and Fig. 7.

Similarly, a synthetic polynucleotide based on the 5'-region of clone 36 was used to screen and isolate cDNAs from the lambda gt-11 NANBV cDNA library which overlapped clone 36 cDNA (Section IV.A.8.). A purified clone of recombinant phage-containing cDNA which hybridized to the synthetic polynucleotide probe was named clone 35 and the NANBV cDNA sequence contained within this clone is shown in Fig. 8.

By utilizing the technique of isolating overlapping cDNA sequences, clones containing additional upstream and downstream HCV cDNA sequences have been obtained. The isolation of these clones, is described infra in Section IV.A.

Analysis of the nucleotide sequences of the HCV cDNAs encoded within the isolated clones show that the composite cDNA contains one long continuous ORF. Fig. 26 shows the sequence of the composite cDNA from these clones, along with the putative HCV polypeptide encoded therein.

The description of the method to retrieve the cDNA sequences is mostly of historical interest. The resultant sequences (and their complements) are provided herein, and the sequences, or any portion thereof, could be prepared using synthetic methods, or by a combination of synthetic methods with retrieval of partial sequences using methods similar to those described herein.

Lambda-gt11 strains replicated from the HCV cDNA library and from clones 5-1-1, 81, 1-2 and 91 have been deposited under the terms of the Budapest Treaty with the American Type Culture Collection (ATCC), 12301 Parklawn

Dr., Rockville, Maryland 20852, and have been assigned the following Accession Numbers.

lambda-gt11	ATCC No.	Deposit Date
HCV cDNA library	40394	1 Dec. 1987
clone 81	40388	17 Nov. 1987
clone 91	40389	17 Nov. 1987
clone 1-2	40390	17 Nov. 1987
clone 5-1-1	40391	18 Nov. 1987

Upon allowance and issuance of this application as a United States Patent, all restriction on availability of these deposits will be irrevocably removed; and access to the designated deposits will be available during pendency of the above-named application to one determined by the Commissioner to be entitled thereto under 37 CFR 1.14 and 35 USC 1.22. Moreover, the designated deposits will be maintained for a period of thirty (30) years from the date of deposit, or for five (5) years after the last request for the deposit; or for the enforceable life of the U.S. patent, whichever is longer. These deposits and other deposited materials mentioned herein are intended for convenience only, and are not required to practice the present invention in view of the description here. The HCV cDNA sequences in all of the deposited materials are incorporated herein by reference.

The description above, of "walking" the genome by isolating overlapping cDNA sequences from the HCV lambda gt-11 library provides one method by which cDNAs corresponding to the entire HCV genome may be isolated. However, given the information provided herein, other methods for isolating these cDNAs are obvious to one of skill in the art. Some of these methods are described in Section IV.A., infra.

II.B. Preparation of Viral Polypeptides and Fragments

The availability of cDNA sequences, either those isolated by utilizing the cDNA sequences in Figs. 1-32, as discussed infra, as well as the cDNA sequences in these figures, permits the construction of expression vectors encoding antigenically active regions of the polypeptide encoded in either strand. These antigenically active regions may be derived from coat or envelope antigens or from core antigens, including, for example, polynucleotide binding proteins, polynucleotide polymerase(s), and other viral proteins required for the replication and/or assembly of the virus particle. Fragments encoding the desired polypeptides are derived from the cDNA clones using conventional restriction digestion or by synthetic methods, and are ligated into vectors which may, for example, contain portions of fusion sequences such as beta-Galactosidase or superoxide dismutase (SOD), preferably SOD. Methods and vectors which are useful for the production of polypeptides which contain fusion sequences of SOD are described in European Patent Office Publication number 0196056, published October 1, 1986. Vectors encoding fusion polypeptides of SOD and HCV polypeptides, i.e., $NANB_{5-1-1}$, $NANB_{81}$, and C100-3, which is encoded in a composite of HCV cDNAs, are described in Sections IV.B.1, IV.B.2, and IV.B.4, respectively. Any desired portion of the HCV cDNA containing an open reading frame, in either sense strand, can be obtained as a recombinant polypeptide, such as a mature or fusion protein; alternatively, a polypeptide encoded in the cDNA can be provided by chemical synthesis.

The DNA encoding the desired polypeptide, whether in fused or mature form, and whether or not containing a signal sequence to permit secretion, may be ligated into expression vectors suitable for any convenient host. Both eukaryotic and prokaryotic host

systems are presently used in forming recombinant polypeptides, and a summary of some of the more common control systems and host cell lines is given in Section III.A., infra. The polypeptide is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use. Purification may be by techniques known in the art, for example, salt fractionation, chromatography on ion exchange resins, affinity chromatography, centrifugation, and the like. See, for example, Methods in Enzymology for a variety of methods for purifying proteins. Such polypeptides can be used as diagnostics, or those which give rise to neutralizing antibodies may be formulated into vaccines. Antibodies raised against these polypeptides can also be used as diagnostics, or for passive immunotherapy. In addition, as discussed in Section II.J. herein below, antibodies to these polypeptides are useful for isolating and identifying HCV particles.

The HCV antigens may also be isolated from HCV virions. The virions may be grown in HCV infected cells in tissue culture, or in an infected host.

II.C. Preparation of Antigenic Polypeptides and Conjugation with Carrier

An antigenic region of a polypeptide is generally relatively small--typically 8 to 10 amino acids or less in length. Fragments of as few as 5 amino acids may characterize an antigenic region. These segments may correspond to regions of HCV antigen. Accordingly, using the cDNAs of HCV as a basis, DNAs encoding short segments of HCV polypeptides can be expressed recombinantly either as fusion proteins, or as isolated polypeptides. In addition, short amino acid sequences can be conveniently obtained by chemical synthesis. In instances wherein the synthesized polypeptide is correctly configured so as to

provide the correct epitope, but is too small to be immunogenic, the polypeptide may be linked to a suitable carrier.

A number of techniques for obtaining such linkage are known in the art, including the formation of disulfide linkages using N-succinimidyl-3-(2-pyridylthio)propionate (SPDP) and succinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate (SMCC) obtained from Pierce Company, Rockford, Illinois, (if the peptide lacks a sulfhydryl group, this can be provided by addition of a cysteine residue.) These reagents create a disulfide linkage between themselves and peptide cysteine residues on one protein and an amide linkage through the epsilon-amino on a lysine, or other free amino group in the other. A variety of such disulfide/amide-forming agents are known. See, for example, Immun. Rev. (1982) 62:185. Other bifunctional coupling agents form a thioether rather than a disulfide linkage. Many of these thio-ether-forming agents are commercially available and include reactive esters of 6-maleimidocaproic acid, 2-bromoacetic acid, 2-iodoacetic acid, 4-(N-maleimidomethyl)cyclohexane-1-carboxylic acid, and the like. The carboxyl groups can be activated by combining them with succinimide or 1-hydroxyl-2-nitro-4-sulfonic acid, sodium salt. The foregoing list is not meant to be exhaustive, and modifications of the named compounds can clearly be used.

Any carrier may be used which does not itself induce the production of antibodies harmful to the host. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins; polysaccharides, such as latex functionalized sepharose, agarose, cellulose, cellulose beads and the like; polymeric amino acids, such as polyglutamic acid, polylysine, and the like; amino acid copolymers; and inactive virus particles, see, for

example, section II.D. Especially useful protein substrates are serum albumins, keyhole limpet hemocyanin, immunoglobulin molecules, thyroglobulin, ovalbumin, tetanus toxoid, and other proteins well known to those skilled in the art.

II.D. <u>Preparation of Hybrid Particle Immunogens Containing HCV Epitopes</u>

The immunogenicity of the epitopes of HCV may also be enhanced by preparing them in mammalian or yeast systems fused with or assembled with particle-forming proteins such as, for example, that associated with hepatitis B surface antigen. Constructs wherein the NANBV epitope is linked directly to the particle-forming protein coding sequences produce hybrids which are immunogenic with respect to the HCV epitope. In addition, all of the vectors prepared include epitopes specific to HBV, having various degrees of immunogenicity, such as, for example, the pre-S peptide. Thus, particles constructed from particle forming protein which include HCV sequences are immunogenic with respect to HCV and HBV.

Hepatitis surface antigen (HBSAg) has been shown to be formed and assembled into particles in <u>S. cerevisiae</u> (Valenzuela et al. (1982)), as well as in, for example, mammalian cells (Valenzuela, P., et al. (1984)). The formation of such particles has been shown to enhance the immunogenicity of the monomer subunit. The constructs may also include the immunodominant epitope of HBSAg, comprising the 55 amino acids of the presurface (pre-S) region. Neurath et al. (1984). Constructs of the pre-S-HBSAg particle expressible in yeast are disclosed in EPO 174,444, published March 19, 1986; hybrids including heterologous viral sequences for yeast expression are disclosed in EPO 175,261, published March 26, 1966. Both applications are assigned to the herein assignee, and are

incorporated herein by reference. These constructs may also be expressed in mammalian cells such as Chinese hamster ovary (CHO) cells using an SV40-dihydrofolate reductase vector (Michelle et al. (1984)).

In addition, portions of the particle-forming protein coding sequence may be replaced with codons encoding an HCV epitope. In this replacement, regions which are not required to mediate the aggregation of the units to form immunogenic particles in yeast or mammals can be deleted, thus eliminating additional HBV antigenic sites from competition with the HCV epitope.

II.E. Preparation of Vaccines

Vaccines may be prepared from one or more immunogenic polypeptides derived from HCV cDNA as well as from the cDNA sequences in the Figs. 1-32, or from the HCV genome to which they correspond. The observed homology between HCV and Flaviviruses provides information concerning the polypeptides which are likely to be most effective as vaccines, as well as the regions of the genome in which they are encoded. The general structure of the Flavivirus genome is discussed in Rice et al (1986). The flavivirus genomic RNA is believed to be the only virus-specific mRNA species, and it is translated into the three viral structural proteins, i.e., C, M, and E, as well as two large nonstructural proteins, NV4 and NV5, and a complex set of smaller nonstructural proteins. It is known that major neutralizing epitopes for Flaviviruses reside in the E (envelope) protein (Roehrig (1986)). The corresponding HCV E gene and polypeptide encoding region can be predicted, based upon the homology to Flaviviruses. Thus, vaccines may be comprised of recombinant polypeptides containing epitopes of HCV E. These polypeptides may be expressed in bacteria, yeast, or mammalian cells, or alternatively may be isolated from

viral preparations. It is also anticipated that the other structural proteins may also contain epitopes which give rise to protective anti-HCV antibodies. Thus, polypeptides containing the epitopes of E, C, and M may also be used, whether singly or in combination, in HCV vaccines.

In addition to the above, it has been shown that immunization with NS1 (nonstructural protein 1), results in protection against yellow fever (Schlesinger et al (1986)). This is true even though the immunization does not give rise to neutralizing antibodies. Thus, particularly since this protein appears to be highly conserved among Flaviviruses, it is likely that HCV NS1 will also be protective against HCV infection. Moreover, it also shows that nonstructural proteins may provide protection against viral pathogenicity, even if they do not cause the production of neutralizing antibodies.

In view of the above, multivalent vaccines against HCV may be comprised of one or more structural proteins, and/or one or more nonstructural proteins. These vaccines may be comprised of, for example, recombinant HCV polypeptides and/or polypeptides isolated from the virions. In addition, it may be possible to use inactivated HCV in vaccines; inactivation may be by the preparation of viral lysates, or by other means known in the art to cause inactivation of Flaviviruses, for example, treatment with organic solvents or detergents, or treatment with formalin. Moreover, vaccines may also be prepared from attenuated HCV strains. The preparation of attenuated HCV strains is described infra.

It is known that some of the proteins in Flaviviruses contain highly conserved regions, thus, some immunological cross-reactivity is expected between HCV and other Flaviviruses. It is possible that shared epitopes between the Flaviviruses and HCV will give rise to

protective antibodies against one or more of the disorders caused by these pathogenic agents. Thus, it may be possible to design multipurpose vaccines based upon this knowledge.

The preparation of vaccines which contain an immunogenic polypeptide(s) as active ingredients, is known to one skilled in the art. Typically, such vaccines are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection may also be prepared. The preparation may also be emulsified, or the protein encapsulated in liposomes. The active immunogenic ingredients are often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine. Examples of adjuvants which may be effective include but are not limited to: aluminum hydroxide, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween 80 emulsion. The effectiveness of an adjuvant may be determined by measuring the amount of antibodies directed against an immunogenic polypeptide containing an HCV antigenic sequence resulting from administration of this polypeptide

in vaccines which are also comprised of the various adjuvants.

The vaccines are conventionally administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, oral formulations. For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1%-2%. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain 10%-95% of active ingredient, preferably 25%-70%.

The proteins may be formulated into the vaccine as neutral or salt forms. Pharmaceutically acceptable salts include the acid addition salts (formed with free amino groups of the peptide) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids such as acetic, oxalic, tartaric, maleic, and the like. Salts formed with the free carboxyl groups may also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

II.F. Dosage and Administration of Vaccines

The vaccines are administered in a manner compatible with the dosage formulation, and in such amount

as will be prophylactically and/or therapeutically effective. The quantity to be administered, which is generally in the range of 5 micrograms to 250 micrograms of antigen per dose, depends on the subject to be treated, capacity of the subject's immune system to synthesize antibodies, and the degree of protection desired. Precise amounts of active ingredient required to be administered may depend on the judgment of the practitioner and may be peculiar to each subject.

The vaccine may be given in a single dose schedule, or preferably in a multiple dose schedule. A multiple dose schedule is one in which a primary course of vaccination may be with 1-10 separate doses, followed by other doses given at subsequent time intervals required to maintain and or reenforce the immune response, for example, at 1-4 months for a second dose, and if needed, a subsequent dose(s) after several months. The dosage regimen will also, at least in part, be determined by the need of the individual and be dependent upon the judgment of the practitioner.

In addition, the vaccine containing the immunogenic HCV antigen(s) may be administered in conjunction with other immunoregulatory agents, for example, immune globulins.

II.G. <u>Preparation of Antibodies Against HCV Epitopes</u>

The immunogenic polypeptides prepared as described above are used to produce antibodies, both polyclonal and monoclonal. If polyclonal antibodies are desired, a selected mammal (e.g., mouse, rabbit, goat, horse, etc.) is immunized with an immunogenic polypeptide bearing an HCV epitope(s). Serum from the immunized animal is collected and treated according to known procedures. If serum containing polyclonal antibodies to an HCV epitope contains antibodies to other antigens, the

polyclonal antibodies can be purified by immunoaffinity chromatography. Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and Walker (1987).

Alternatively, polyclonal antibodies may be isolated from a mammal which has been previously infected with HCV. An example of a method for purifying antibodies to HCV epitopes from serum from an infected individual, based upon affinity chromatography and utilizing a fusion polypeptide of SOD and a polypeptide encoded within cDNA clone 5-1-1, is presented in Section V.E.

Monoclonal antibodies directed against HCV epitopes can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies by hybridomas is well known. Immortal antibody-producing cell lines can be created by cell fusion, and also by other techniques such as direct transformation of B lymphocytes with oncogenic DNA, or transfection with Epstein-Barr virus. See, e.g., M. Schreier et al. (1980); Hammerling et al. (1981); Kennett et al. (1980); see also, U.S. Patent Nos. 4,341,761; 4,399,121; 4,427,783; 4,444,887; 4,466,917; 4,472,500; 4,491,632; and 4,493,890. Panels of monoclonal antibodies produced against HCV epitopes can be screened for various properties; i.e., for isotype, epitope affinity, etc.

Antibodies, both monoclonal and polyclonal, which are directed against HCV epitopes are particularly useful in diagnosis, and those which are neutralizing are useful in passive immunotherapy. Monoclonal antibodies, in particular, may be used to raise anti-idiotype antibodies.

Anti-idiotype antibodies are immunoglobulins which carry an "internal image" of the antigen of the infectious agent against which protection is desired.

See, for example, Nisonoff, A., et al. (1981) and Dreesman et al. (1985).

Techniques for raising anti-idiotype antibodies are known in the art. See, for example, Grzych (1985), MacNamara et al. (1984), and Uytdehaag et al. (1985). These anti-idiotype antibodies may also be useful for treatment of NANBH, as well as for an elucidation of the immunogenic regions of HCV antigens.

II.H. Diagnostic Oligonucleotide Probes and Kits

Using the disclosed portions of the isolated HCV cDNAs as a basis, including those in Figs. 1-32, oligomers of approximately 8 nucleotides or more can be prepared, either by excision or synthetically, which hybridize with the HCV genome and are useful in identification of the viral agent(s), further characterization of the viral genome(s), as well as in detection of the virus(es) in diseased individuals. The probes for HCV polynucleotides (natural or derived) are a length which allows the detection of unique viral sequences by hybridization. While 6-8 nucleotides may be a workable length, sequences of 10-12 nucleotides are preferred, and about 20 nucleotides appears optimal. Preferably, these sequences will derive from regions which lack heterogeneity. These probes can be prepared using routine methods, including automated oligonucleotide synthetic methods. Among useful probes, for example, are the clone 5-1-1 and the additional clones disclosed herein, as well as the various oligomers useful in probing cDNA libraries, set forth below. A complement to any unique portion of the HCV genome will be satisfactory. For use as probes, complete complementarity is desirable, though it may be unnecessary as the length of the fragment is increased.

For use of such probes as diagnostics, the biological sample to be analyzed, such as blood or serum, is

treated, if desired, to extract the nucleic acids contained therein. The resulting nucleic acid from the sample may be subjected to gel electrophoresis or other size separation techniques; alternatively, the nucleic acid sample may be dot blotted without size separation. The probes are then labeled. Suitable labels, and methods for labeling probes are known in the art, and include, for example, radioactive labels incorporated by nick translation or kinasing, biotin, fluorescent probes, and chemiluminescent probes. The nucleic acids extracted from the sample are then treated with the labeled probe under hybridization conditions of suitable stringencies.

The probes can be made completely complementary to the HCV genome. Therefore, usually high stringency conditions are desirable in order to prevent false positives. However, conditions of high stringency should only be used if the probes are complementary to regions of the viral genome which lack heterogeneity. The stringency of hybridization is determined by a number of factors during hybridization and during the washing procedure, including temperature, ionic strength, length of time, and concentration of formamide. These factors are outlined in, for example, Maniatis, T. (1982).

Generally, it is expected that the HCV genome sequences will be present in serum of infected individuals at relatively low levels, i.e., at approximately 10^2-10^3 sequences per ml. This level may require that amplification techniques be used in hybridization assays. Such techniques are known in the art. For example, the Enzo Biochemical Corporation "Bio-Bridge" system uses terminal deoxynucleotide transferase to add unmodified 3'-poly-dT-tails to a DNA probe. The poly dT-tailed probe is hybridized to the target nucleotide sequence, and then to a biotin-modified poly-A. PCT application 84/03520 and EPA124221 describe a DNA hybridization assay in which: (1)

analyte is annealed to a single-stranded DNA probe that is complementary to an enzyme-labeled oligonucleotide; and (2) the resulting tailed duplex is hybridized to an enzyme-labeled oligonucleotide. EPA 204510 describes a DNA hybridization assay in which analyte DNA is contacted with a probe that has a tail, such as a poly-dT tail, an amplifier strand that has a sequence that hybridizes to the tail of the probe, such as a poly-A sequence, and which is capable of binding a plurality of labeled strands. A particularly desirable technique may first involve amplification of the target HCV sequences in sera approximately 10,000 fold, i.e., to approximately 10^6 sequences/ml. This may be accomplished, for example, by the technique of Saiki et al. (1986). The amplified sequence(s) may then be detected using a hybridization assay which is described in co-pending U.S. Application, Attorney Docket No. 2300-0171, which was filed 15 October 1987, is assigned to the herein assignee, and is hereby incorporated herein by reference. This hybridization assay, which should detect sequences at the level of 10^6/ml utilizes nucleic acid multimers which bind to single-stranded analyte nucleic acid, and which also bind to a multiplicity of single-stranded labeled oligonucleotides. A suitable solution phase sandwich assay which may be used with labeled polynucleotide probes, and the methods for the preparation of probes is described in EPO 225,807, published June 16, 1987, which is assigned to the herein assignee, and which is hereby incorporated herein by reference.

The probes can be packaged into diagnostic kits. Diagnostic kits include the probe DNA, which may be labeled; alternatively, the probe DNA may be unlabeled and the ingredients for labeling may be included in the kit. The kit may also contain other suitably packaged reagents and materials needed for the particular hybridization

protocol, for example, standards, as well as instructions for conducting the test.

II.I. Immunoassay and Diagnostic Kits

Both the polypeptides which react immunologically with serum containing HCV antibodies, for example, those derived from or encoded within the clones described in Section IV.A., and composites thereof, (see section IV.A.) and the antibodies raised against the HCV specific epitopes in these polypeptides, see for example Section IV.E, are useful in immunoassays to detect presence of HCV antibodies, or the presence of the virus and/or viral antigens, in biological samples, including for example, blood or serum samples. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. For example, the immunoassay may utilize one viral antigen, for example, a polypeptide derived from any of the clones containing HCV cDNA described in Section IV.A., or from the composite cDNAs derived from the cDNAs in these clones, or from the HCV genome from which the cDNA in these clones is derived; alternatively, the immunoassay may use a combination of viral antigens derived from these sources. It may use, for example, a monoclonal antibody directed towards a viral epitope(s), a combination of monoclonal antibodies directed towards one viral antigen, monoclonal antibodies directed towards different viral antigens, polyclonal antibodies directed towards the same viral antigen, or polyclonal antibodies directed towards different viral antigens. Protocols may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive,

or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

The Flavivirus model for HCV allows predictions regarding the likely location of diagnostic epitopes for the virion structural proteins. The C, pre-M, M, and E domains are all likely to contain epitopes of significant potential for detecting viral antigens, and particularly for diagnosis. Similarly, domains of the nonstructural proteins are expected to contain important diagnostic epitopes (e.g., NS5 encoding a putative polymerase; and NS1 encoding a putative complement-binding antigen). Recombinant polypeptides, or viral polypeptides, which include epitopes from these specific domains may be useful for the detection of viral antibodies in infections blood donors and infected patients.

In addition, antibodies directed against the E and/or M proteins can be used in immunoassays for the detection of viral antigens in patients with HCV caused NANBH, and in infectious blood donors. Moreover, these antibodies will be extremely useful in detecting acute-phase donors and patients.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the polypeptides of the invention containing HCV epitopes or antibodies directed against HCV epitopes in suitable containers, along with the remaining reagents and materials required for the conduct of the assay, as well as a suitable set of assay instructions.

II.J. Further Characterization of the HCV Genome, Virions, and Viral Antigens Using Probes Derived From cDNA to the Viral Genome

The HCV cDNA sequence information in the clones described in Section IV.A., as shown in Figs. 1-32, inclusive, may be used to gain further information on the sequence of the HCV genome, and for identification and isolation of the HCV agent, and thus will aid in its characterization including the nature of the genome, the structure of the viral particle, and the nature of the antigens of which it is composed. This information, in turn, can lead to additional polynucleotide probes, polypeptides derived from the HCV genome, and antibodies directed against HCV epitopes which would be useful for the diagnosis and/or treatment of HCV caused NANBH.

The cDNA sequence information in the above-mentioned clones is useful for the design of probes for the isolation of additional cDNA sequences which are derived from as yet undefined regions of the HCV genome(s) from which the cDNAs in clones described in Section IV.A. are derived. For example, labeled probes containing a sequence of approximately 8 or more nucleotides, and preferably 20 or more nucleotides, which are derived from regions close to the 5'-termini or 3'-termini of the family of HCV cDNA sequences shown in Figs. 1, 3, 6, 9, 14 and 32 may be used to isolate overlapping cDNA sequences from HCV cDNA libraries. These sequences which overlap the cDNAs in the above-mentioned clones, but which also contain sequences derived from regions of the genome from which the cDNA in the above mentioned clones are not derived, may then be used to synthesize probes for identification of other overlapping fragments which do not necessarily overlap the cDNAs in the clones described in Section IV.A. Unless the HCV genome is segmented and the segments lack common sequences, it is possible to sequence

the entire viral genome(s) utilizing the technique of isolation of overlapping cDNAs derived from the viral genome(s). Although it is unlikely, if the genome is a segmented genome which lacks common sequences, the sequence of the genome can be determined by serologically screening lambda-gt11 HCV cDNA libraries, as used to isolate clone 5-1-1, sequencing cDNA isolates, and using the isolated cDNAs to isolate overlapping fragments, using the technique described for the isolation and sequencing of the clones described in Section IV.A. Alternatively, characterization of the genomic segments could be from the viral genome(s) isolated from purified HCV particles. Methods for purifying HCV particles and for detecting them during the purification procedure are described herein, infra. Procedures for isolating polynucleotide genomes from viral particles are known in the art, and one procedure which may be used is shown in Example IV.A.1. The isolated genomic segments could then be cloned and sequenced. Thus, with the information provided herein, it is possible to clone and sequence the HCV genome(s) irrespective of their nature.

Methods for constructing cDNA libraries are known in the art, and are discussed supra and infra; a method for the construction of HCV cDNA libraries in lambda-gt11 is discussed infra in Section IV.A. However, cDNA libraries which are useful for screening with nucleic acid probes may also be constructed in other vectors known in the art, for example, lambda-gt10 (Huynh et al. (1985)). The HCV derived cDNA detected by the probes derived from the cDNAs in Figs. 1-32, and from the probes synthesized from polynucleotides derived from these cDNAs, may be isolated from the clone by digestion of the isolated polynucleotide with the appropriate restriction enzyme(s), and sequenced. See, for example, Section IV.A.3. and IV.A.4. for the techniques used for the

isolation and sequencing of HCV cDNA which overlaps HCV cDNA in clone 5-1-1, Sections IV.A.5-IV.A.7 for the isolation and sequencing of HCV cDNA which overlaps that in clone 81, and Section IV.A.8 and IV.A.9 for the isolation and sequencing of a clone which overlaps another clone (clone 36), which overlaps clone 81.

The sequence information derived from these overlapping HCV cDNAs is useful for determining areas of homology and heterogeneity within the viral genome(s), which could indicate the presence of different strains of the genome, and/or of populations of defective particles. It is also useful for the design of hybridization probes to detect HCV or HCV antigens or HCV nucleic acids in biological samples, and during the isolation of HCV (discussed infra), utilizing the techniques described in Section II.G. Moreover, the overlapping cDNAs may be used to create expression vectors for polypeptides derived from the HCV genome(s) which also encode the polypeptides encoded in clones 5-1-1, 36, 81, 91, and 1-2, and in the other clones described in Section IV.A. The techniques for the creation of these polypeptides containing HCV epitopes, and for antibodies directed against HCV epitopes contained within them, as well as their uses, are analogous to those described for polypeptides derived from NANBV cDNA sequences contained within clones 5-1-1, 32, 35, 36, 1-2, 81, and 91, discussed supra and infra.

Encoded within the family of cDNA sequences contained within clones 5-1-1, 32, 35, 36, 81, 91, 1-2, and the other clones described in Section IV.A. are antigen(s) containing epitopes which appear to be unique to HCV; i.e., antibodies directed against these antigens are absent from individuals infected with HAV or HBV, and from individuals not infected with HCV (see the serological data presented in Section IV.B.). Moreover, a comparison of the sequence information of these cDNAs with

the sequences of HAV, HBV, HDV, and with the genomic sequences in Genebank indicates that minimal homology exists between these cDNAs and the polynucleotide sequences of those sources. Thus, antibodies directed against the antigens encoded within the cDNAs of these clones may be used to identify BB-NANBV particles isolated from infected individuals. In addition, they are also useful for the isolation of NANBH agent(s).

HCV particles may be isolated from the sera from BB-NANBV infected individuals or from cell cultures by any of the methods known in the art, including for example, techniques based on size discrimination such as sedimentation or exclusion methods, or techniques based on density such as ultracentrifugation in density gradients, or precipitation with agents such as polyethylene glycol, or chromatography on a variety of materials such as anionic or cationic exchange materials, and materials which bind due to hydrophobicity, as well as affinity columns. During the isolation procedure the presence of HCV may be detected by hybridization analysis of the extracted genome, using probes derived from the HCV cDNAs described supra, or by immunoassay (see Section II.I.) utilizing as probes antibodies directed against HCV antigens encoded within the family of cDNA sequences shown in Figs. 1-32, and also directed against HCV antigens encoded within the overlapping HCV cDNA sequences discussed supra. The antibodies may be monoclonal, or polyclonal, and it may be desirable to purify the antibodies before their use in the immunoassay. A purification procedure for polyclonal antibodies directed against antigen(s) encoded within clone 5-1-1 is described in Section IV.E; analogous purification procedures may be utilized for antibodies directed against other HCV antigens.

Antibodies directed against HCV antigens encoded within the family of cDNAs shown in Figs. 1-32, as well as

those encoded within overlapping HCV cDNAs, which are affixed to solid supports are useful for the isolation of HCV by immunoaffinity chromatography. Techniques for immunoaffinity chromatography are known in the art, including techniques for affixing antibodies to solid supports so that they retain their immunoselective activity; the techniques may be those in which the antibodies are adsorbed to the support (see, for example, Kurstak in ENZYME IMMUNODIAGNOSIS, page 31-37), as well as those in which the antibodies are covalently linked to the support. Generally, the techniques are similar to those used for covalent linking of antigens to a solid support, which are generally described in Section II.C.; however, spacer groups may be included in the bifunctional coupling agents so that the antigen binding site of the antibody remains accessible.

During the purification procedure the presence of HCV may be detected and/or verified by nucleic acid hybridization, utilizing as probes polynucleotides derived from the family of HCV cDNA sequences shown in Figs. 1-32, as well as from overlapping HCV cDNA sequences, described supra. In this case, the fractions are treated under conditions which would cause the disruption of viral particles, for example, with detergents in the presence of chelating agents, and the presence of viral nucleic acid determined by hybridization techniques described in Section II.H. Further confirmation that the isolated particles are the agents which induce HCV may be obtained by infecting chimpanzees with the isolated virus particles, followed by a determination of whether the symptoms of NANBH result from the infection.

Viral particles from the purified preparations may then be further characterized. The genomic nucleic acid has been purified. Based upon its sensitivity to RNase, and not DNase I, it appears that the virus is

composed of an RNA genome. See Example IV.C.2., infra. The strandedness and circularity or non-circularity can determined by techniques known in the art, including, for example, its visualization by electron microscopy, its migration in density gradients, and its sedimentation characteristics. Based upon the hybridization of the captured HCV genome to the negative strands of HCV cDNAs, it appears that HCV may be comprised of a positive stranded RNA genome (see Section IV.H.1). Techniques such as these are described in, for example, METHODS IN ENZYMOLOGY. In addition, the purified nucleic acid can be cloned and sequenced by known techniques, including reverse transcription since the genomic material is RNA. See, for example, Maniatis (1982), and Glover (1985). Utilizing the nucleic acid derived from the viral particles, it is possible to sequence the entire genome, whether or not it is segmented.

Examination of the homology of the polypeptide encoded within the continuous ORF of combined clones 14i through 39c (see Fig. 26), shows that the HCV polypeptide contains regions of homology with the corresponding proteins in conserved regions of flaviviruses. An example of this is described in Section IV.H.3. This finding has many important ramifications. First, this evidence, in conjunction with the results which show that HCV contains a positive-stranded genome, the size of which is approximately 10,000 nucleotides, is consistent with the suggestion that HCV is a flavivirus, or flavi-like virus. Generally, flavivirus virions and their genomes have a relatively consistent structure and organization, which are known. See Rice et al. (1986), and Brinton, M.A. (1988). Thus, the structural genes encoding the polypeptides C, pre-M/M, and E may be located in the 5'-terminus of the genome upstream of clone 14i. Moreover, using the comparison with other flaviviruses, predictions

as to the precise location of the sequences encoding these proteins can be made.

Isolation of the sequences upstream of those in clone 14i may be accomplished in a number of ways which, given the information herein, would be obvious to one of skill in the art. For example, the genome "walking" technique, may be used to isolate other sequences which are 5′ to those in clone 14i, but which overlap that clone; this in turn leads to the isolation of additional sequences. This technique has been amply demonstrated infra, in Section IV.A.. For example, also, it is known that the flaviviruses have conserved epitopes and regions of conserved nucleic acid sequences. Polynucleotides containing the conserved sequences may be used as probes which bind the HCV genome, thus allowing its isolation. In addition, these conserved sequences, in conjunction with those derived from the HCV cDNAs shown in Fig. 22, may be used to design primers for use in systems which amplify the genome sequences upstream of those in clone 14i, using polymerase chain reaction technology. An example of this is described infra.

The structure of the HCV may also be determined and its components isolated. The morphology and size may be determined by, for example, electron microscopy. The identification and localization of specific viral polypeptide antigens such as coat or envelope antigens, or internal antigens, such as nucleic acid binding proteins, core antigens, and polynucleotide polymerase(s) may also be determined by, for example, determining whether the antigens are present as major or minor viral components, as well as by utilizing antibodies directed against the specific antigens encoded within isolated cDNAs as probes. This information is useful in the design of vaccines; for example, it may be preferable to include an exterior antigen in a vaccine preparation. Multivalent vaccines

may be comprised of, for example, a polypeptide derived from the genome encoding a structural protein, for example, E, as well as a polypeptide from another portion of the genome, for example, a nonstructural or structural polypeptide.

II.K. Cell Culture Systems and Animal Model Systems for HCV Replication

The suggestion that HCV is a flavivirus or flavi-like virus also provides information on methods for growing HCV. The term "flavi-like" means that the virus shows a significant amount of homology to the known conserved regions of flaviviruses and that the majority of the genome is a single ORF. Methods for culturing flaviviruses are known to those of skill in the art (See, for example, the reviews by Brinton (1986) and Stollar, V. (1980)). Generally, suitable cells or cell lines for culturing HCV may include those known to support Flavivirus replication, for example, the following: monkey kidney cell lines (e.g. MK_2, VERO); porcine kidney cell lines (e.g. PS); baby hamster kidney cell lines (e.g. BHK); murine macrophage cell lines (e.g., P388D1, MK1, Mm1); human macrophage cell lines (e.g., U-937); human peripheral blood leukocytes; human adherent monocytes; hepatocytes or hepatocyte cell lines (e.g., HUH7 , HEPG2); embryos or embryonic cells (e.g., chick embryo fibroblasts); or cell lines derived from invertebrates, preferably from insects (e.g. drosophila cell lines), or more preferably from arthropods, for example, mosquito cell lines (e.g., A. Albopictus, Aedes aegypti, Cutex tritaeniorhynchus) or tick cell lines (e.g. RML-14 Dermacentor parumapertus).

It is possible that primary hepatocytes can be cultured, and then infected with HCV; or alternatively, the hepatocyte cultures could be derived from the livers

of infected individuals (e.g., humans or chimpanzees). The latter case is an example of a cell which is infected in vivo being passaged in vitro. In addition, various immortalization methods can be used to obtain cell-lines derived from hepatocyte cultures. For example, primary liver cultures (before and after enrichment of the hepatocyte population) may be fused to a variety of cells to maintain stability. For example, also, cultures may be infected with transforming viruses, or transfected with transforming genes in order to create permanent or semipermanent cell lines. In addition, for example, cells in liver cultures may be fused to established cell lines (e.g., HepG2). Methods for cell fusion are known in the art, and include, for example, the use of fusion agents such as polyethylene glycol, Sendai Virus, and Epstein-Barr virus.

As discussed above, HCV is a Flavivirus or Flavi-like virus. Therefore, it is probable that HCV infection of cell lines may be accomplished by techniques known in the art for infecting cells with Flaviviruses. These include, for example, incubating the cells with viral preparations under conditions which allow viral entry into the cell. In addition, it may be possible to obtain viral production by transfecting the cells with isolated viral polynucleotides. It is known that Togavirus and Flavivirus RNAs are infectious in a variety of vertebrate cell lines (Pfefferkorn and Shapiro (1974)), and in a mosquito cell line (Peleg (1969)). Methods for transfecting tissue culture cells with RNA duplexes, positive stranded RNAs, and DNAs (including cDNAs) are known in the art, and include, for example, techniques which use electroporation, and precipitation with DEAE-Dextran or calcium phosphate. An abundant source of HCV RNA can be obtained by performing in vitro transcription of an HCV cDNA corresponding to the complete

genome. Transfection with this material, or with cloned HCV cDNA should result in viral replication and the *in vitro* propagation of the virus.

In addition to cultured cells, animal model systems may be used for viral replication; animal systems in which flaviviruses are known to those of skill in the art (See, for example, the review by Monath (1986)). Thus, HCV replication may occur not only in chimpanzees, but also in, for example, marmosets and suckling mice.

II.L. Screening for Anti-Viral Agents for HCV

The availability of cell culture and animal model systems for HCV also makes possible screening for anti-viral agents which inhibit HCV replication, and particularly for those agents which preferentially allow cell growth and multiplication while inhibiting viral replication. These screening methods are known by those of skill in the art. Generally, the anti-viral agents are tested at a variety of concentrations, for their effect on preventing viral replication in cell culture systems which support viral replication, and then for an inhibition of infectivity or of viral pathogenicity (and a low level of toxicity) in an animal model system.

The methods and compositions provided herein for detecting HCV antigens and HCV polynucleotides are useful for screening of anti-viral agents in that they provide an alternative, and perhaps more sensitive means, for detecting the agent's effect on viral replication than the cell plaque assay or ID_{50} assay. For example, the HCV-polynucleotide probes described herein may be used to quantitate the amount of viral nucleic acid produced in a cell culture. This could be accomplished, for example, by hybridization or competition hybridization of the infected cell nucleic acids with a labeled HCV-polynucleotide probe. For example, also, anti-HCV antibodies may be used

to identify and quantitate HCV antigen(s) in the cell culture utilizing the immunoassays described herein. In addition, since it may be desirable to quantitate HCV antigens in the infected cell culture by a competition assay, the polypeptides encoded within the HCV cDNAs described herein are useful in these competition assays. Generally, a recombinant HCV polypeptide derived from the HCV cDNA would be labeled, and the inhibition of binding of this labeled polypeptide to an HCV polypeptide due to the antigen produced in the cell culture system would be monitored. Moreover, these techniques are particularly useful in cases where the HCV may be able to replicate in a cell line without causing cell death.

II.M. Preparation of Attenuated Strains of HCV

In addition to the above, utilizing the tissue culture systems and/or animal model systems, it may be possible to isolate attenuated strains of HCV. These strains would be suitable for vaccines, or for the isolation of viral antigens. Attenuated strains are isolatable after multiple passages in cell culture and/or an animal model. Detection of an attenuated strain in an infected cell or individual is achievable by techniques known in the art, and could include, for example, the use of antibodies to one or more epitopes encoded in HCV as a probe or the use of a polynucleotide containing an HCV sequence of at least about 8 nucleotides as a probe. Alternatively, or in addition, an attenuated strain may be constructed utilizing the genomic information of HCV provided herein, and utilizing recombinant techniques. Generally, one would attempt to delete a region of the genome encoding, for example, a polypeptide related to pathogenicity, but which allows viral replication. In addition, the genome construction would allow the expression of an epitope which gives rise to neutralizing

antibodies for HCV. The altered genome could then be utilized to transform cells which allow HCV replication, and the cells grown under conditions to allow viral replication. Attenuated HCV strains are useful not only for vaccine purposes, but also as sources for the commercial production of viral antigens, since the processing of these viruses would require less stringent protection measures for the employees involved in viral production and/or the production of viral products.

III. General Methods

The general techniques used in extracting the genome from a virus, preparing and probing a cDNA library, sequencing clones, constructing expression vectors, transforming cells, performing immunological assays such as radioimmunoassays and ELISA assays, for growing cells in culture, and the like are known in the art and laboratory manuals are available describing these techniques. However, as a general guide, the following sets forth some sources currently available for such procedures, and for materials useful in carrying them out.

III.A. Hosts and Expression Control Sequences

Both prokaryotic and eukaryotic host cells may be used for expression of desired coding sequences when appropriate control sequences which are compatible with the designated host are used. Among prokaryotic hosts, *E. coli* is most frequently used. Expression control sequences for prokaryotes include promoters, optionally containing operator portions, and ribosome binding sites. Transfer vectors compatible with prokaryotic hosts are commonly derived from, for example, pBR322, a plasmid containing operons conferring ampicillin and tetracycline resistance, and the various pUC vectors, which also contain sequences conferring antibiotic resistance mark-

ers. These markers may be used to obtain successful transformants by selection. Commonly used prokaryotic control sequences include the Beta-lactamase (penicillinase) and lactose promoter systems (Chang et al. (1977)), the tryptophan (trp) promoter system (Goeddel et al. (1980)) and the lambda-derived P_L promoter and N gene ribosome binding site (Shimatake et al. (1981)) and the hybrid _tac_ promoter (De Boer et al. (1983)) derived from sequences of the _trp_ and _lac_ UV5 promoters. The foregoing systems are particularly compatible with _E. coli_; if desired, other prokaryotic hosts such as strains of Bacillus or Pseudomonas may be used, with corresponding control sequences.

Eukaryotic hosts include yeast and mammalian cells in culture systems. _Saccharomyces cerevisiae_ and _Saccharomyces carlsbergensis_ are the most commonly used yeast hosts, and are convenient fungal hosts. Yeast compatible vectors carry markers which permit selection of successful transformants by conferring prototrophy to auxotrophic mutants or resistance to heavy metals on wild-type strains. Yeast compatible vectors may employ the 2 micron origin of replication (Broach et al. (1983)), the combination of CEN3 and ARS1 or other means for assuring replication, such as sequences which will result in incorporation of an appropriate fragment into the host cell genome. Control sequences for yeast vectors are known in the art and include promoters for the synthesis of glycolytic enzymes (Hess et al. (1968); Holland et al. (1978)), including the promoter for 3 phosphoglycerate kinase (Hitzeman (1980)). Terminators may also be included, such as those derived from the enolase gene (Holland (1981)). Particularly useful control systems are those which comprise the glyceraldehyde-3 phosphate dehydrogenase (GAPDH) promoter or alcohol dehydrogenase (ADH) regulatable promoter, terminators also derived from

GAPDH, and if secretion is desired, leader sequence from yeast alpha factor. In addition, the transcriptional regulatory region and the transcriptional initiation region which are operably linked may be such that they are not naturally associated in the wild-type organism. These systems are described in detail in EPO 120,551, published October 3, 1984; EPO 116,201, published August 22, 1984; and EPO 164,556, published December 18, 1985, all of which are assigned to the herein assignee, and are hereby incorporated herein by reference.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including HeLa cells, Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells, and a number of other cell lines. Suitable promoters for mammalian cells are also known in the art and include viral promoters such as that from Simian Virus 40 (SV40) (Fiers (1978)), Rous sarcoma virus (RSV), adenovirus (ADV), and bovine papilloma virus (BPV). Mammalian cells may also require terminator sequences and poly A addition sequences; enhancer sequences which increase expression may also be included, and sequences which cause amplification of the gene may also be desirable. These sequences are known in the art. Vectors suitable for replication in mammalian cells may include viral replicons, or sequences which insure integration of the appropriate sequences encoding NANBV epitopes into the host genome.

III.B. Transformations

Transformation may be by any known method for introducing polynucleotides into a host cell, including, for example packaging the polynucleotide in a virus and transducing a host cell with the virus, and by direct uptake of the polynucleotide. The transformation

procedure used depends upon the host to be transformed. For example, transformation of the E. coli host cells with lambda-gt11 containing BB-NANBV sequences is discussed in the Example section, infra. Bacterial transformation by direct uptake generally employs treatment with calcium or rubidium chloride (Cohen (1972); Maniatis (1982)). Yeast transformation by direct uptake may be carried out using the method of Hinnen et al. (1978). Mammalian transformations by direct uptake may be conducted using the calcium phosphate precipitation method of Graham and Van der Eb (1978), or the various known modifications thereof.

III.C. Vector Construction

Vector construction employs techniques which are known in the art. Site-specific DNA cleavage is performed by treating with suitable restriction enzymes under conditions which generally are specified by the manufacturer of these commercially available enzymes. In general, about 1 microgram of plasmid or DNA sequence is cleaved by 1 unit of enzyme in about 20 microliters buffer solution by incubation of 1-2 hr at 37° C. After incubation with the restriction enzyme, protein is removed by phenol/chloroform extraction and the DNA recovered by precipitation with ethanol. The cleaved fragments may be separated using polyacrylamide or agarose gel electrophoresis techniques, according to the general procedures found in Methods in Enzymology (1980) 65:499-560.

Sticky ended cleavage fragments may be blunt ended using E. coli DNA polymerase I (Klenow) in the presence of the appropriate deoxynucleotide triphosphates (dNTPs) present in the mixture. Treatment with S1 nuclease may also be used, resulting in the hydrolysis of any single stranded DNA portions.

Ligations are carried out using standard buffer and temperature conditions using T4 DNA ligase and ATP;

sticky end ligations require less ATP and less ligase than blunt end ligations. When vector fragments are used as part of a ligation mixture, the vector fragment is often treated with bacterial alkaline phosphatase (BAP) or calf intestinal alkaline phosphatase to remove the 5'-phosphate and thus prevent religation of the vector; alternatively, restriction enzyme digestion of unwanted fragments can be used to prevent ligation.

Ligation mixtures are transformed into suitable cloning hosts, such as E. coli, and successful transformants selected by, for example, antibiotic resistance, and screened for the correct construction.

III.D. Construction of Desired DNA Sequences

Synthetic oligonucleotides may be prepared using an automated oligonucleotide synthesizer as described by Warner (1984). If desired the synthetic strands may be labeled with ^{32}P by treatment with polynucleotide kinase in the presence of ^{32}P-ATP, using standard conditions for the reaction.

DNA sequences, including those isolated from cDNA libraries, may be modified by known techniques, including, for example site directed mutagenesis, as described by Zoller (1982). Briefly, the DNA to be modified is packaged into phage as a single stranded sequence, and converted to a double stranded DNA with DNA polymerase using, as a primer, a synthetic oligonucleotide complementary to the portion of the DNA to be modified, and having the desired modification included in its own sequence. The resulting double stranded DNA is transformed into a phage supporting host bacterium. Cultures of the transformed bacteria, which contain replications of each strand of the phage, are plated in agar to obtain plaques. Theoretically, 50% of the new plaques contain phage having the mutated sequence, and the

remaining 50% have the original sequence. Replicates of the plaques are hybridized to labeled synthetic probe at temperatures and conditions which permit hybridization with the correct strand, but not with the unmodified sequence. The sequences which have been identified by hybridization are recovered and cloned.

III.E. Hybridization with Probe

DNA libraries may be probed using the procedure of Grunstein and Hogness (1975). Briefly, in this procedure, the DNA to be probed is immobilized on nitrocellulose filters, denatured, and prehybridized with a buffer containing 0-50% formamide, 0.75 M NaCl, 75 mM Na citrate, 0.02% (wt/v) each of bovine serum albumin, polyvinyl pyrollidone, and Ficoll, 50 mM Na Phosphate (pH 6.5), 0.1% SDS, and 100 micrograms/ml carrier denatured DNA. The percentage of formamide in the buffer, as well as the time and temperature conditions of the prehybridization and subsequent hybridization steps depends on the stringency required. Oligomeric probes which require lower stringency conditions are generally used with low percentages of formamide, lower temperatures, and longer hybridization times. Probes containing more than 30 or 40 nucleotides such as those derived from cDNA or genomic sequences generally employ higher temperatures, e.g., about 40-42^{o}C, and a high percentage, e.g., 50%, formamide. Following prehybridization, 5'-^{32}P-labeled oligonucleotide probe is added to the buffer, and the filters are incubated in this mixture under hybridization conditions. After washing, the treated filters are subjected to autoradiography to show the location of the hybridized probe; DNA in corresponding locations on the original agar plates is used as the source of the desired DNA.

III.F. Verification of Construction and Sequencing

For routine vector constructions, ligation mixtures are transformed into *E. coli* strain HB101 or other suitable host, and successful transformants selected by antibiotic resistance or other markers. Plasmids from the transformants are then prepared according to the method of Clewell et al. (1969), usually following chloramphenicol amplification (Clewell (1972)). The DNA is isolated and analyzed, usually by restriction enzyme analysis and/or sequencing. Sequencing may be by the dideoxy method of Sanger et al. (1977) as further described by Messing et al. (1981), or by the method of Maxam et al. (1980). Problems with band compression, which are sometimes observed in GC rich regions, were overcome by use of T-deazoguanosine according to Barr et al. (1986).

III.G. Enzyme Linked Immunosorbent Assay

The enzyme-linked immunosorbent assay (ELISA) can be used to measure either antigen or antibody concentrations. This method depends upon conjugation of an enzyme to either an antigen or an antibody, and uses the bound enzyme activity as a quantitative label. To measure antibody, the known antigen is fixed to a solid phase (e.g., a microplate or plastic cup), incubated with test serum dilutions, washed, incubated with anti-immunoglobulin labeled with an enzyme, and washed again. Enzymes suitable for labeling are known in the art, and include, for example, horseradish peroxidase. Enzyme activity bound to the solid phase is measured by adding the specific substrate, and determining product formation or substrate utilization colorimetrically. The enzyme activity bound is a direct function of the amount of antibody bound.

To measure antigen, a known specific antibody is fixed to the solid phase, the test material containing antigen is added, after an incubation the solid phase is washed, and a second enzyme-labeled antibody is added. After washing, substrate is added, and enzyme activity is estimated colorimetrically, and related to antigen concentration.

IV. Examples

Described below are examples of the present invention which are provided only for illustrative purposes, and not to limit the scope of the present invention. In light of the present disclosure, numerous embodiments within the scope of the claims will be apparent to those of ordinary skill in the art. The procedures set forth, for example, in Sections IV.A. may, if desired, be repeated but need not be, as techniques are available for construction of the desired nucleotide sequences based on the information provided by the invention. Expression is exemplified in *E. coli*; however, other systems are available as set forth more fully in Section III.A. Additional epitopes derived from the genomic structure may also be produced, and used to generate antibodies as set forth below.

IV.A. Preparation, Isolation and Sequencing of HCV cDNA

IV.A.1. Preparation of HCV cDNA

The source of NANB agent was a plasma pool derived from a chimpanzee with chronic NANBH. The chimpanzee had been experimentally infected with blood from another chimpanzee with chronic NANBH resulting from infection with HCV in a contaminated batch of factor 8 concentrate derived from pooled human sera. The chimpanzee plasma pool was made by combining many

individual plasma samples containing high levels of alanine aminotransferase activity; this activity results from hepatic injury due to the HCV infection. Since 1 ml of a 10^{-6} dilution of this pooled serum given i.v. caused NANBH in another chimpanzee, its CID was at least 10^6/ml, i.e., it had a high infectious virus titer.

A cDNA library from the high titer plasma pool was generated as follows. First, viral particles were isolated from the plasma; a 90 ml aliquot was diluted with 310 ml of a solution containing 50 mM Tris-HCl, pH 8.0, 1mM EDTA, 100 mM NaCl. Debris was removed by centrifugation for 20 min at 15,000 x g at 20°C. Viral particles in the resulting supernatant were then pelleted by centrifugation in a Beckman SW28 rotor at 28,000 rpm for 5 hours at 20°C. To release the viral genome, the particles were disrupted by suspending the pellets in 15 ml solution containing 1% sodium dodecyl sulfate (SDS), 10 mM EDTA, 10 mM Tris-HCl, pH 7.5, also containing 2 mg/ml proteinase k, followed by incubation at 45°C for 90 min. Nucleic acids were isolated by adding 0.8 micrograms MS2 bacteriophage RNA as carrier, and extracting the mixture four times with a 1:1 mixture of phenol:chloroform (phenol saturated with 0.5M Tris-HCl, pH 7.5, 0.1% (v/v) beta-mercaptoethanol, 0.1% (w/v) hydroxyquinolone, followed by extraction two times with chloroform. The aqueous phase was concentrated with 1-butanol prior to precipitation with 2.5 volumes absolute ethanol overnight at -20°C. Nucleic acid was recovered by centrifugation in a Beckman SW41 rotor at 40,000 rpm for 90 min at 4°C, and dissolved in water that had been treated with 0.05% (v/v) diethylpyrocarbonate and autoclaved.

Nucleic acid obtained by the above procedure (<2 micrograms) was denatured with 17.5 mM CH_3HgOH; cDNA was synthesized using this denatured nucleic acid as template, and was cloned into the EcoRI site of phage lambda-gt11

using methods described by Huynh (1985), except that random primers replaced oligo(dT) 12-18 during the synthesis of the first cDNA strand by reverse transcriptase (Taylor et al. (1976)). The resulting double stranded cDNAs were fractionated according to size on a Sepharose CL-4B column; eluted material of approximate mean size 400, 300, 200, and 100 base-pairs were pooled into cDNA pools 1, 2, 3, and 4, respectively. The lambda-gt11 cDNA library was generated from the cDNA in pool 3.

The lambda-gt11 cDNA library generated from pool 3 was screened for epitopes that could bind specifically with serum derived from a patient who had previously experienced NANBH. About 10^6 phage were screened with patient sera using the methods of Huynh et al. (1985), except that bound human antibody was detected with sheep anti-human Ig antisera that had been radio-labeled with ^{125}I. Five positive phages were identified and purified. The five positive phages were then tested for specificity of binding to sera from 8 different humans previously infected with the NANBH agent, using the same method. Four of the phage encoded a polypeptide that reacted immunologically with only one human serum, i.e., the one that was used for primary screening of the phage library. The fifth phage (5-1-1) encoded a polypeptide that reacted immunologically with 5 of 8 of the sera tested. Moreover, this polypeptide did not react immunologically with sera from 7 normal blood donors. Therefore, it appears that clone 5-1-1 encodes a polypeptide which is specifically recognized immunologically by sera from NANB patients.

IV.A.2. <u>Sequences of the HCV cDNA in Recombinant Phage 5-1-1, and of the Polypeptide Encoded Within the Sequence.</u>

The cDNA in recombinant phage 5-1-1 was sequenced by the method of Sanger et al. (1977). Es-

sentially, the cDNA was excised with EcoRI, isolated by size fractionation using gel electrophoresis. The EcoRI restriction fragments were subcloned into the M13 vectors, mp18 and mp19 (Messing (1983)) and sequenced using the dideoxychain termination method of Sanger et al. (1977). The sequence obtained is shown in Fig. 1.

The polypeptide encoded in Fig. 1 that is encoded in the HCV cDNA is in the same translational frame as the N-terminal beta-galactosidase moiety to which it is fused. As shown in Section IV.A., the translational open reading frame (ORF) of 5-1-1 encodes epitope(s) specifically recognized by sera from patients and chimpanzees with NANBH infections.

IV.A.3. <u>Isolation of Overlapping HCV cDNA to cDNA in Clone 5-1-1.</u>

Overlapping HCV cDNA to the cDNA in clone 5-1-1 was obtained by screening the same lambda-gt11 library, created as described in Section IV.A.1., with a synthetic polynucleotide derived from the sequence of the HCV cDNA in clones 5-1-1, as shown in Fig. 1. The sequence of the polynucleotide used for screening was:

5'-TCC CTT GCT CGA TGT ACG GTA AGT GCT GAG AGC ACT CTT CCA TCT CAT CGA ACT CTC GGT AGA GGA CTT CCC TGT CAG GT-3'.

The lambda-gt11 library was screened with this probe, using the method described in Huynh (1985). Approximately 1 in 50,000 clones hybridized with the probe. Three clones which contained cDNAs which hybridized with the synthetic probe have been numbered 81, 1-2, and 91.

IV.A.4. Nucleotide Sequences of Overlapping HCV cDNAs to cDNA in Clone 5-1-1.

The nucleotide sequences of the three cDNAs in clones 81, 1-2, and 91 were determined essentially as in Section IV.A.2. The sequences of these clones relative to the HCV cDNA sequence in phage 5-1-1 is shown in Fig. 2, which shows the strand encoding the detected HCV epitope, and where the homologies in the nucleotide sequences are indicated by vertical lines between the sequences.

The sequences of the cloned HCV cDNAs are highly homologous in the overlapping regions (see Fig. 2). However, there are differences in two regions. Nucleotide 67 in clone 1-2 is a thymidine, whereas the other three clones contain a cytidine residue in this position. It should be noted, however, that the same amino acid is encoded when either C or T occupies this position.

The second difference is that clone 5-1-1 contains 28 base pairs which are not present in the other three clones. These base pairs occur at the start of the cDNA sequence in 5-1-1, and are indicated by small letters. Based on radioimmunoassay data, which is discussed infra in Section IV.D., it is possible that an HCV epitope may be encoded in this 28 bp region.

The absence of the 28 base pairs of 5-1-1 from clones 81, 1-2, and 91 may mean that the cDNA in these clones were derived from defective HCV genomes; alternatively, the 28 bp region could be a terminal artifact in clone 5-1-1.

The sequences of small letters in the nucleotide sequence of clones 81 and 91 simply indicate that these sequences have not been found in other cDNAs because cDNAs overlapping these regions were not yet isolated.

A composite HCV cDNA sequence derived from overlapping cDNAs in clones 5-1-1, 81, 1-2 and 91 is shown in Fig. 3. However, in this figure the unique 28 base pairs

of clone 5-1-1 are omitted. The figure also shows the sequence of the polypeptide encoded within the ORF of the composite HCV cDNA.

IV.A.5. Isolation of Overlapping HCV cDNAs to cDNA in Clone 81.

The isolation of HCV cDNA sequences upstream of, and which overlap those in clone 81 cDNA was accomplished as follows. The lambda-gt11 cDNA library prepared as described in Section IV.A.1. was screened by hybridization with a synthetic polynucleotide probe which was homologous to a 5' terminal sequence of clone 81. The sequence of clone 81 is presented in Fig. 4. The sequence of the synthetic polynucleotide used for screening was:

5' CTG TCA GGT ATG ATT GCC GGC TTC CCG GAC 3'.

The methods were essentially as described in Huynh (1985), except that the library filters were given two washes under stringent conditions, i.e., the washes were in 5 x SSC, 0.1% SDS at 55°C for 30 minutes each. Approximately 1 in 50,000 clones hybridized with the probe. A positive recombinant phage which contained cDNA which hybridized with the sequence was isolated and purified. This phage has been numbered clone 36.

Downstream cDNA sequences, which overlaps the carboxyl-end sequences in clone 81 cDNA were isolated using a procedure similar to that for the isolation of upstream cDNA sequences, except that a synthetic oligonucleotide probe was prepared which is homologous to a 3' terminal sequence of clone 81. The sequence of the synthetic polynucleotide used for screening was:

5' TTT GGC TAG TGG TTA GTG GGC TGG TGA CAG 3'

A positive recombinant phage, which contained cDNA which hybridized with this latter sequence was isolated and purified, and has been numbered clone 32.

IV.A.6. Nucleotide Sequence of HCV cDNA in Clone 36.

The nucleotide sequence of the cDNA in clone 36 was determined essentially as described in Section IV.A.2. The double-stranded sequence of this cDNA, its region of overlap with the HCV cDNA in clone 81, and the polypeptide encoded by the ORF are shown in Fig. 5.

The ORF in clone 36 is in the same translational frame as the HCV antigen encoded in clone 81. Thus, in combination, the ORFs in clones 36 and 81 encode a polypeptide that represents part of a large HCV antigen. The sequence of this putative HCV polypeptide and the double stranded DNA sequence encoding it, which is derived from the combined ORFs of the HCV cDNAs of clones 36 and 81, is shown in Fig. 6.

IV.A.7 Nucleotide Sequences of HCV cDNA in Clone 32

The nucleotide sequence of the cDNA in clone 32 was determined essentially as was that described in Section IV.A.2 for the sequence of clone 5-1-1. The sequence data indicated that the cDNA in clone 32 recombinant phage was derived from two different sources. One fragment of the cDNA was comprised of 418 nucleotides derived from the HCV genome; the other fragment was comprised of 172 nucleotides derived from the bacteriophage MS2 genome, which had been used as a carrier during the preparation of the lambda gt11 plasma cDNA library.

The sequence of the cDNA in clone 32 corresponding to that of the HCV genome is shown in Fig. 7. The region of the sequences that overlaps that of clone 81, and the polypeptide encoded by the ORF are also indicated in the figure. This sequence contains one continuous ORF

that is in the same translational frame as the HCV antigen encoded by clone 81.

IV.A.8 Isolation of Overlapping HCV cDNA to cDNA in Clone 36

The isolation of HCV cDNA sequences upstream of, and which overlap those in clone 36 cDNA was accomplished as described in Section IV.A.5, for those which overlap clone 81 cDNA, except that the synthetic polynucleotide was based on the 5'-region of clone 36. The sequence of the synthetic polynucleotide used for screening was:

5' AAG CCA CCG TGT GCG CTA GGG CTC AAG CCC 3'

Approximately 1 in 50,000 clones hybridized with the probe. The isolated, purified clone of recombinant phage which contained cDNA which hybridized to this sequence was named clone 35.

IV.A.9 Nucleotide Sequence of HCV cDNA in Clone 35

The nucleotide sequence of the cDNA in clone 35 was determined essentially as described in Section IV.A.2. The sequence, its region of overlap with that of the cDNA in clone 36, and the putative polypeptide encoded therein, are shown in Fig. 8.

Clone 35 apparently contains a single, continuous ORF that encodes a polypeptide in the same translational frame as that encoded by clone 36, clone 81, and clone 32. Fig. 9 shows the sequence of the long continuous ORF that extends through clones 35, 36, 81, and 32, along with the putative HCV polypeptide encoded therein. This combined sequence has been confirmed using other independent cDNA clones derived from the same lambda gt11 cDNA library.

IV.A.10. Isolation of Overlapping HCV cDNA to cDNA in Clone 35

The isolation of HCV cDNA sequences upstream of, and which overlap those in clone 35 cDNA was accomplished as described in Section IV.A.8, for those which overlap clone 36 cDNA, except that the synthetic polynucleotide was based on the 5'-region of clone 35. The sequence of the synthetic polynucleotide used for screening was:

5' CAG GAT GCT GTC TCC CGC ACT CAA CGT 3'

Approximately 1 in 50,000 clones hybridized with the probe. The isolated, purified clone of recombinant phage which contained cDNA which hybridized to this sequence was named clone 37b.

IV.A.11. Nucleotide Sequence of HCV in Clone 37b

The nucleotide sequence of the cDNA in clone 37b was determined essentially as described in Section IV.A.2. The sequence, its region of overlap with that of the cDNA in clone 35, and the putative polypeptide encoded therein, are shown in Fig. 10.

The 5'-terminal nucleotide of clone 35 is a T, whereas the corresponding nucleotide in clone 37b is an A. The cDNAs from three other independent clones which were isolated during the procedure in which clone 37b was isolated, described in Section IV.A.10, have also been sequenced. The cDNAs from these clones also contain an A in this position. Thus, the 5'-terminal T in clone 35 may be an artefact of the cloning procedure. It is known that artefacts often arise at the 5'-termini of cDNA molecules.

Clone 37b apparently contains one continuous ORF which encodes a polypeptide which is a continuation of the polypeptide encoded in the ORF which extends through the overlapping clones 35, 36, 81 and 32.

IV.A.12 <u>Isolation of Overlapping HCV cDNA to cDNA in Clone 32</u>

The isolation of HCV cDNA sequences downstream of clone 32 was accomplished as follows. First, clone cla was isolated utilizing a synthetic hybridization probe which was based on the nucleotide sequence of the HCV cDNA sequence in clone 32. The method was essentially that described in Section IV.A.5, except that the sequence of the synthetic probe was:

5′ AGT GCA GTG GAT GAA CCG GCT GAT AGC CTT 3′.

Utilizing the nucleotide sequence from clone cla, another synthetic nucleotide was synthesized which had the sequence:

5′ TCC TGA GGC GAC TGC ACC AGT GGA TAA GCT 3′.

Screening of the lambda gt11 library using the clone cla derived sequence as probe yielded approximately 1 in 50,000 positive colonies. An isolated, purified clone which hybridized with this probe was named clone 33b.

IV.A.13 <u>Nucleotide Sequence of HCV cDNA in Clone 33b</u>

The nucleotide sequence of the cDNA in clone 33b was determined essentially as described in Section IV.A.2. The sequence, its region of overlap with that of the cDNA in clone 32, and the putative polypeptide encoded therein, are shown in Fig. 11.

Clone 33b apparently contains one continuous ORF which is an extension of the ORFs in overlapping clones 37b, 35, 36, 81 and 32. The polypeptide encoded in clone 33b is in the same translational frame as that encoded in the extended ORF of these overlapping clones.

IV.A.14 Isolation of Overlapping HCV cDNAs to cDNA Clone 37b and to cDNA in Clone 33b

In order to isolate HCV cDNAs which overlap the cDNAs in clone 37b and in clone 33b, the following synthetic oligonucleotide probes, which were derived from the cDNAs in those clones, were used to screen the lambda gt11 library, using essentially the method described in Section IV.A.3. The probes used were:

5′ CAG GAT GCT GTC TCC CGC ACT CAA CGT C 3′

and

5′ TCC TGA GGC GAC TGC ACC AGT GGA TAA GCT 3′

to detect colonies containing HCV cDNA sequences which overlap those in clones 37b and 33b, respectively. Approximately 1 in 50,000 colonies were detected with each probe. A clone which contained cDNA which was upstream of, and which overlapped the cDNA in clone 37b, was named clone 40b. A clone which contained cDNA which was downstream of, and which overlapped the cDNA in clone 33b was named clone 25c.

IV.A.15 Nucleotide Sequences of HCV cDNA in clone 40b and in clone 25c

The nucleotide sequences of the cDNAs in clone 40b and in clone 25c were determined essentially as described in Section IV.A.2. The sequences of 40b and 25c, their regions of overlap with the cDNAs in clones 37b and 33b, and the putative polypeptides encoded therein, are shown in Fig. 12 (clone 40b) and Fig. 13 (clone 25c).

The 5′-terminal nucleotide of clone 40b is a G. However, the cDNAs from five other independent clones

which were isolated during the procedure in which clone 40b was isolated, described in Section IV.A.14, have also been sequenced. The cDNAs from these clones also contain a T in this position. Thus, the G may represent a cloning artifact (see the discussion in Section IV.A.11).

The 5′-terminus of clone 25c is ACT, but the sequence of this region in clone cla (sequence not shown), and in clone 33b is TCA. This difference may also represent a cloning artifact, as may the 28 extra 5′-terminal nucleotides in clone 5-1-1.

Clones 40b and 25c each apparently contain an ORF which is an extension of the continuous ORF in the previously sequenced clones. The nucleotide sequence of the ORF extending through clones 40b, 37b, 35, 36, 81, 32, 33b, and 25c, and the amino acid sequence of the putative polypeptide encoded therein, are shown in Fig. 14. In the figure, the potential artifacts have been omitted from the sequence, and instead, the corresponding sequences in non-5′-terminal regions of multiple overlapping clones are shown.

IV.A.16. Preparation of a Composite HCV cDNA from the cDNAs in Clones 36, 81, and 32

The composite HCV cDNA, C100, was constructed as follows. First the cDNAs from the clones 36, 81, and 32 were excised with EcoRI. The EcoRI fragment of cDNA from each clone was cloned individually into the EcoRI site of the vector pGEM3-blue (Promega Biotec). The resulting recombinant vectors which contained the cDNAs from clones 36, 81, and 32 were named pGEM3-blue/36, pGEM3-blue/81, and pGEM3-blue/32, respectively. The appropriately oriented recombinant of pGEM3-blue/81 was digested with NaeI and NarI, and the large (~2850bp) fragment was purified and ligated with the small (~570bp) NaeI/NarI purified restriction fragment from pGEM3-blue/36. This

composite of the cDNAs from clones 36 and 81 was used to generate another pGEM3-blue vector containing the continuous HCV ORF contained within the overlapping cDNA within these clones. This new plasmid was then digested with PvuII and EcoRI to release a fragment of approximately 680bp, which was then ligated with the small (580bp) PvuII/EcoRI fragment isolated from the appropriately oriented pGEM3-blue/32 plasmid, and the composite cDNA from clones 36, 81, and 32 was ligated into the EcoRI linearized vector pSODcf1, which is described in Section IV.B.1, and which was used to express clone 5-1-1 in bacteria. Recombinants containing the ~1270bp EcoRI fragment of composite HCV cDNA (C100) were selected, and the cDNA from the plasmids was excised with EcoRI and purified.

IV.A.17. Isolation and Nucleotide Sequences of HCV cDNAs in Clones 14i, 11b, 7f, 7e, 8h, 33c, 14c, 8f, 33f, 33g, and 39c

The HCV cDNAs in clones 14i, 11b, 7f, 7e, 8h, 33c, 14c, 8f, 33f, 33g, and 39c were isolated by the technique of isolating overlapping cDNA fragments from the lambda gt11 library of HCV cDNAs described in Section IV.A.1.. The technique used was essentially as described in Section IV.A.3., except that the probes used were designed from the nucleotide sequence of the last isolated clones from the 5′ and the 3′ end of the combined HCV sequences. The frequency of clones which hybridized with the probes described below was approximately 1 in 50,000 in each case.

The nucleotide sequences of the HCV cDNAs in clones 14i, 7f, 7e, 8h, 33c, 14c, 8f, 33f, 33g, and 39c were determined essentially as described in Section IV.A.2., except that the cDNA excised from these phages were substituted for the cDNA isolated from clone 5-1-1.

Clone 33c was isolated using a hybridization probe based on the sequence of nucleotides in clone 40b. The nucleotide sequence of clone 40b is presented in Fig. 12. The nucleotide sequence of the probe used to isolate 33c was:

5′ ATC AGG ACC GGG GTG AGA ACA ATT ACC ACT 3′

The sequence of the HCV cDNA in clone 33c, and the overlap with that in clone 40b, is shown in Fig. 15, which also shows the amino acids encoded therein.

Clone 8h was isolated using a probe based on the sequence of nucleotides in clone 33c. The nucleotide sequence of the probe was

5′ AGA GAC AAC CAT GAG GTC CCC GGT GTT C 3′.

The sequence of the HCV cDNA in clone 8h, and the overlap with that in clone 33c, and the amino acids encoded therein, are shown in Fig. 16.

Clone 7e was isolated using a probe based on the sequence of nucleotides in clone 8h. The nucleotide sequence of the probe was

5′ TCG GAC CTT TAC CTG GTC ACG AGG CAC 3′.

The sequence of HCV cDNA in clone 7e, the overlap with clone 8h, and the amino acids encoded therein, are shown in Fig. 17.

Clone 14c was isolated with a probe based on the sequence of nucleotides in clone 25c. The sequence of clone 25c is shown in Fig. 13. The probe in the isolation of clone 14c had the sequence

5′ ACC TTC CCC ATT AAT GCC TAC ACC ACG GGC 3′.

The sequence of HCV cDNA in clone 14c, its overlap with that in clone 25c, and the amino acids encoded therein are shown in Fig. 18.

Clone 8f was isolated using a probe based on the sequence of nucleotides in clone 14c. The nucleotide sequence of the probe was

5′ TCC ATC TCT CAA GGC AAC TTG CAC CGC TAA 3′.

The sequence of HCV cDNA in clone 8f, its overlap with that in clone 14c, and the amino acids encoded therein are shown in Fig. 19.

Clone 33f was isolated using a probe based on the nucleotide sequence present in clone 8f. The nucleotide sequence of the probe was

5′ TCC ATG GCT GTC CGC TTC CAC CTC CAA AGT 3′.

The sequence of HCV cDNA in clone 33f, its overlap with that in clone 8f, and the amino acids encoded therein are shown in Fig. 20.

Clone 33g was isolated using a probe based on the sequence of nucleotides in clone 33f. The nucleotide sequence of the probe was

5′ GCG ACA ATA CGA CAA CAT CCT CTG AGC CCG 3′.

The sequence of HCV cDNA in clone 33g, its overlap with that in clone 33f, and the amino acids encoded therein are shown in Fig. 21.

Clone 7f was isolated using a probe based on the sequence of nucleotides in clone 7e. The nucleotide sequence of the probe was

5′ AGC AGA CAA GGG GCC TCC TAG GGT GCA TAA T 3′.

The sequence of HCV cDNA in clone 7f, its overlap with clone 7e, and the amino acids encoded therein are shown in Fig. 22.

Clone 11b was isolated using a probe based on the sequence of clone 7f. The nucleotide sequence of the probe was

5′ CAC CTA TGT TTA TAA CCA TCT CAC TCC TCT 3′.

The sequence of HCV cDNA in clone 11b, its overlap with clone 7f, and the amino acids encoded therein are shown in Fig. 23.

Clone 14i was isolated using a probe based on the sequence of nucleotides in clone 11b. The nucleotide sequence of the probe was

5′ CTC TGT CAC CAT ATT ACA AGC GCT ATA TCA 3′.

The sequence of HCV cDNA in clone 14i, its overlap with 11b, and the amino acids encoded therein are shown in Fig. 24.

Clone 39c was isolated using a probe based on the sequence of nucleotides in clone 33g. The nucleotide sequence of the probe was

5′ CTC GTT GCT ACG TCA CCA CAA TTT GGT GTA 3′

The sequence of HCV cDNA in clone 39c, its overlap with clone 33g, and the amino acids encoded therein are shown in Fig. 25.

IV.A.18. <u>The Composite HCV cDNA Sequence Derived from Isolated Clones Containing HCV cDNA</u>

The HCV cDNA sequences in the isolated clones described supra have been aligned to create a composite

HCV cDNA sequence. The isolated clones, aligned in the 5′ to 3′ direction are: 14i, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g, and 39c.

A composite HCV cDNA sequence derived from the isolated clones, and the amino acids encoded therein, is shown in Fig. 26.

In creating the composite sequence the following sequence heterogeneities have been considered. Clone 33c contains an HCV cDNA of 800 base pairs, which overlaps the cDNAs in clones 40b and 37c. In clone 33c, as well as in 5 other overlapping clones, nucleotide #789 is a G. However, in clone 37b (see Section IV.A.11), the corresponding nucleotide is an A. This sequence difference creates an apparent heterogeneity in the amino acids encoded therein, which would be either CYS or TYR, for G or A, respectively. This heterogeneity may have important ramifications in terms of protein folding.

Nucleotide residue #2 in clone 8h HCV cDNA is a T. However, as shown infra, the corresponding residue in clone 7e is an A; moreover, an A in this position is also found in 3 other isolated overlapping clones. Thus, the T residue in clone 8h may represent a cloning artifact. Therefore, in Fig. 26, the residue in this position is designated as an A.

The 3′-terminal nucleotide in clone 8f HCV cDNA is a G. However, the corresponding residue in clone 33f, and in 2 other overlapping clones is a T. Therefore, in Fig. 26, the residue in this position is designated as a T.

The 3′-terminal sequence in clone 33f HCV cDNA is TTGC. However, the corresponding sequence in clone 33g and in 2 other overlapping clones is ATTC. Therefore, in Fig. 26, the corresponding region is represented as ATTC.

Nucleotide residue #4 in clone 33g HCV cDNA is a T. However, in clone 33f and in 2 other overlapping

clones the corresponding residue is an A. Therefore, in Fig. 26, the corresponding residue is designated as an A.

The 3'-terminus of clone 14i is an AA, whereas the corresponding dinucleotide in clone 11b, and in three other clones, is TA. Therefore, in Fig. 26, the TA residue is depicted.

The resolution of other sequence heterogeneities is discussed supra.

An examination of the composite HCV cDNA indicates that it contains one large ORF. This suggests that the viral genome is translated into a large polypeptide which is processed concomitant with, or subsequent to translation.

IV.A.19. <u>Isolation and Nucleotide Sequences of HCV cDNAs in Clones 12f, 35f, 19g, 26g, and 15e</u>

The HCV cDNAs in clones 12f, 35f, 19g, 26g, and 15e were isolated essentially by the technique described in Section IV.A.17, except that the probes were as indicated below. The frequency of clones which hybridized with the probes was approximately 1 in 50,000 in each case. The nucleotide sequences of the HCV cDNAs in these clones were determined essentially as described in Section IV.A.2., except that the cDNA from the indicated clones were substituted for the cDNA isolated from clone 5-1-1.

The isolation of clone 12f, which contains cDNA upstream of the HCV cDNA in Fig. 26, was accomplished using a hybridization probe based on the sequence of nucleotides in clone 14i. The nucleotide sequence of the probe was

5' TGC TTG TGG ATG ATG CTA CTC ATA TCC CAA 3'.

The HCV cDNA sequence of clone 12f, its overlap with clone 14i, and the amino acids encoded therein are shown in Fig. 27.

The isolation of clone 35f, which contains cDNA downstream of the HCV cDNA in Fig. 26, was accomplished using a hybridization probe based on the sequence of nucleotides in clone 39c. The nucleotide sequence of the probe was

5′ AGC AGC GGC GTC AAA AGT GAA GGC TAA CTT 3′.

The sequence of clone 35f, its overlap with the sequence in clone 39c, and the amino acids encoded therein are shown in Fig. 28.

The isolation of clone 19g was accomplished using a hybridization probe based on the 3′ sequence of clone 35f. The nucleotide sequence of the probe was

5′ TTC TCG TAT GAT ACC CGC TGC TTT GAC TCC 3′.

The HCV cDNA sequence of clone 19g, its overlap with the sequence in clone 35f, and the amino acids encoded therein are shown in Fig. 29.

The isolation of clone 26g was accomplished using a hybridization probe based on the 3′ sequence of clone 19g. The nucleotide sequence of the probe was

5′ TGT GTG GCG ACG ACT TAG TCG TTA TCT GTG 3′.

The HCV cDNA sequence of clone 26g, its overlap with the sequence in clone 19g, and the amino acids encoded therein are shown in Fig. 30.

Clone 15e was isolated using a hybridization probe based on the 3′ sequence of clone 26 g. The nucleotide sequence of the probe was

5′ CAC ACT CCA GTC AAT TCC TGG CTA GGC AAC 3′.

The HCV cDNA sequence of clone 15e, its overlap with the sequence in clone 26g, and the amino acids encoded therein are shown in Fig. 31.

The clones described in this Section have been deposited with the ATCC under the terms and conditions described in Section II.A., and have been assigned the following Accession Numbers.

lambda-gt11	ATCC No.	Deposit Date
clone 12f	40514	10 Nov. 1988
clone 35f	40511	10 Nov. 1988
clone 15e	40513	10 Nov. 1988
clone K9-1	40512	10 Nov. 1988

The HCV cDNA sequences in the isolated clones described supra. have been aligned to create a composite HCV cDNA sequence. The isolated clones, aligned in the 5′ to 3′ direction are: 12f, 14i, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f 33f, 33g, 39c, 35f, 19g, 26g, and 15e.

A composite HCV cDNA sequence derived from the isolated clones, and the amino acids encoded therein, is shown in Fig. 32.

IV.A.20. <u>Alternative Method of Isolating cDNA Sequences Upstream of the HCV cDNA Sequence in Clone 12f</u>

Based on the most 5′ HCV sequence in Fig. 32, which is derived from the HCV cDNA in clone 12f, small synthetic oligonucleotide primers of reverse transcriptase are synthesized and used to bind to the corresponding sequence in HCV genomic RNA, to prime reverse transcription of the upstream sequences. The primer

sequences are proximal to the known 5'-terminal sequence of clone 12f, but sufficiently downstream to allow the design of probe sequences upstream of the primer sequences. Known standard methods of priming and cloning are used. The resulting cDNA libraries are screened with sequences upstream of the priming sites (as deduced from the elucidated sequence in clone 12f). The HCV genomic RNA is obtained from either plasma or liver samples from chimpanzees with NANBH, or from analogous samples from humans with NANBH.

IV.A.21. <u>Alternative Method Utilizing Tailing to Isolate Sequences from the 5'-Terminal Region of the HCV Genome</u>

In order to isolate the extreme 5'-terminal sequences of the HCV RNA genome, the cDNA product of the first round of reverse transcription, which is duplexed with the template RNA, is tailed with oligo C. This is accomplished by incubating the product with terminal transferase in the presence of CTP. The second round of cDNA synthesis, which yields the complement of the first strand of cDNA, is accomplished utilizing oligo G as a primer for the reverse transcriptase reaction. The sources of genomic HCV RNA are as described in Section IV.A.20. The methods for tailing with terminal transferase, and for the reverse transcriptase reactions are as in Maniatis et al. (1982). The cDNA products are then cloned, screened, and sequenced.

IV.A.22. <u>Alternative Method Utilizing Tailing to Isolate Sequences from the 3'-Terminal Region of the HCV Genome</u>

This method is based on previously used methods for cloning cDNAs of Flavivirus RNA. In this method, the RNA is subjected to denaturing conditions to remove secondary structures at the 3'-terminus, and is then tailed with Poly A polymerase using rATP as a substrate.

Reverse transcription of the poly A tailed RNA is catalyzed by reverse transcriptase, utilizing oligo dT as a primer. The second strands of cDNA are synthesized, the cDNA products are cloned, screened, and sequenced.

IV.A.23 <u>Creation of Lambda-gt11 HCV cDNA Libraries Containing Larger cDNA Inserts</u>

The method used to create and screen the Lambda gt11 libraries are essentially as described in Section IV.A.1., except that the library is generated from a pool of larger size cDNAs eluted from the Sepharose CL-4B column.

IV.A.24. <u>Creation of HCV cDNA Libraries Using Synthetic Oligomers as Primers</u>

New HCV cDNA libraries have been prepared from the RNA derived from the infectious chimpanzee plasma pool described in Section IV.A.1., and from the poly A^+ RNA fraction derived from the liver of this infected animal. The cDNA was constructed essentially as described by Gubler and Hoffman (1983), except that the primers for the first cDNA strand synthesis were two synthetic oligomers based on the sequence of the HCV genome described supra. Primers based on the sequence of clone 11 b and 7e were, respectively,

5′ CTG GCT TGA AGA ATC 3′

and

5′ AGT TAG GCT GGT GAT TAT GC 3′.

The resulting cDNAs were cloned into lambda bacteriophage vectors, and screened with various other synthetic

oligomers, whose sequences were based on the HCV sequence in Fig. 32.

IV.B. <u>Expression of Polypeptides Encoded Within HCV cDNAs and Identification of the Expressed Products as HCV Induced Antigens</u>.

IV.B.1. <u>Expression of the Polypeptide Encoded in Clone 5-1-1</u>.

The HCV polypeptide encoded within clone 5-1-1 (see Section IV.A.2., supra) was expressed as a fusion polypeptide with superoxide dismutase (SOD). This was accomplished by subcloning the clone 5-1-1 cDNA insert into the expression vector pSODcf1 (Steimer et al. (1986)) as follows.

First, DNA isolated from pSODcf1 was treated with BamHI and EcoRI, and the following linker was ligated into the linear DNA created by the restriction enzymes:

```
5'  GAT CCT GGA ATT CTG ATA A       3'
3'       GA CCT TAA GAC TAT TTT AA  5'
```

After cloning, the plasmid containing the insert was isolated.

Plasmid containing the insert was restricted with EcoRI. The HCV cDNA insert in clone 5-1-1 was excised with EcoRI, and ligated into this EcoRI linearized plasmid DNA. The DNA mixture was used to transform *E. coli* strain D1210 (Sadler et al. (1980)). Recombinants with the 5-1-1 cDNA in the correct orientation for expression of the ORF shown in Fig. 1 were identified by restriction mapping and nucleotide sequencing.

Recombinant bacteria from one clone were induced to express the SOD-NANB$_{5-1-1}$ polypeptide by growing the bacteria in the presence of IPTG.

IV.B.2. Expression of the Polypeptide Encoded in Clone 81.

The HCV cDNA contained within clone 81 was expressed as a SOD-NANB$_{81}$ fusion polypeptide. The method for preparing the vector encoding this fusion polypeptide was analogous to that used for the creation of the vector encoding SOD-NANB$_{5-1-1}$, except that the source of the HCV cDNA was clone 81, which was isolated as described in Section IV.A.3, and for which the cDNA sequence was determined as described in Section IV.A.4. The nucleotide sequence of the HCV cDNA in clone 81, and the putative amino acid sequence of the polypeptide encoded therein are shown in Fig. 4.

The HCV cDNA insert in clone 81 was excised with EcoRI, and ligated into the pSODcf1 which contained the linker (see IV.B.1.) and which was linearized by treatment with EcoRI. The DNA mixture was used to transform *E. coli* strain D1210. Recombinants with the clone 81 HCV cDNA in the correct orientation for expression of the ORF shown in Fig. 4 were identified by restriction mapping and nucleotide sequencing.

Recombinant bacteria from one clone were induced to express the SOD-NANB$_{81}$ polypeptide by growing the bacteria in the presence of IPTG.

IV.B.3. Identification of the Polypeptide Encoded Within Clone 5-1-1 as an HCV and NANBH Associated Antigen.

The polypeptide encoded within the HCV cDNA of clone 5-1-1 was identified as a NANBH associated antigen by demonstrating that sera of chimpanzees and humans infected with NANBH reacted immunologically with the fusion polypeptide, SOD-NANB$_{5-1-1}$, which is comprised of superoxide dismutase at its N-terminus and the in-frame 5-1-1 antigen at its C-terminus. This was accomplished by "Western" blotting (Towbin et al. (1979)) as follows.

A recombinant strain of bacteria transformed with an expression vector encoding the SOD-NANB$_{5-1-1}$ polypeptide, described in Section IV.B.I., was induced to express the fusion polypeptide by growth in the presence of IPTG. Total bacterial lysate was subjected to electrophoresis through polyacrylamide gels in the presence of SDS according to Laemmli (1970). The separated polypeptides were transferred onto nitrocellulose filters (Towbin et al. (1979)). The filters were then cut into thin strips, and the strips were incubated individually with the different chimpanzee and human sera. Bound antibodies were detected by further incubation with ^{125}I-labeled sheep anti-human Ig, as described in Section IV.A.1.

The characterization of the chimpanzee sera used for the Western blots and the results, shown in the photograph of the autoradiographed strips, are presented in Fig. 33. Nitrocellulose strips containing polypeptides were incubated with sera derived from chimpanzees at different times during acute NANBH (Hutchinson strain) infections (lanes 1-16), hepatitis A infections (lanes 17-24, and 26-33), and hepatitis B infections (lanes 34-44). Lanes 25 and 45 show positive controls in which the immunoblots were incubated with serum from the patient used to identify the recombinant clone 5-1-1 in the original screening of the lambda-gt11 cDNA library (see Section IV.A.1.).

The band visible in the control lanes, 25 and 45, in Fig. 23 reflects the binding of antibodies to the NANB$_{5-1-1}$ moiety of the SOD fusion polypeptide. These antibodies do not exhibit binding to SOD alone, since this has also been included as a negative control in these samples, and would have appeared as a band migrating significantly faster than the SOD-NANB$_{5-1-1}$ fusion polypeptide.

Lanes 1-16 of Fig. 33 show the binding of antibodies in sera samples of 4 chimpanzees; the sera were obtained just prior to infection with NANBH, and sequentially during acute infection. As seen from the figure, whereas antibodies which reacted immunologically with the SOD-$NANB_{5-1-1}$ polypeptide were absent in sera samples obtained before administration of infectious HCV inoculum and during the early acute phase of infection, all 4 animals eventually induced circulating antibodies to this polypeptide during the late part of, or following the acute phase. Additional bands observed on the immunoblots in the cases of chimps numbers 3 and 4 were due to background binding to host bacterial proteins.

In contrast to the results obtained with sera from chimps infected with NANBH, the development of antibodies to the $NANB_{5-1-1}$ moiety of the fusion polypeptide was not observed in 4 chimpanzees infected with HAV or 3 chimpanzees infected with HBV. The only binding in these cases was background binding to the host bacterial proteins, which also occurred in the HCV infected samples.

The characterization of the human sera used for the Western blots, and the results, which are shown in the photograph of the autoradiographed strips, are presented in Fig. 34. Nitrocellulose strips containing polypeptides were incubated with sera derived from humans at different times during infections with NANBH (lanes 1-21), HAV (lanes 33-40), and HBV (lanes 41-49). Lanes 25 and 50 show positive controls in which the immunoblots were incubated with serum from patient used in the original screening of the lambda-gt11 library, described supra. Lanes 22-24 and 26-32 show "non-infected" controls in which the sera was from "normal" blood donors.

As seen in Fig. 34, sera from nine NANBH patients, including the serum used for screening the lambda-gt11 library, contained antibodies to the $NANB_{5-1-1}$

moiety of the fusion polypeptide. Sera from three patients with NANBH did not contain these antibodies. It is possible that the anti-NANB$_{5-1-1}$ antibodies will develop at a future date in these patients. It is also possible that this lack of reaction resulted from a different NANBV agent being causative of the disease in the individuals from which the non-responding serum was taken.

Fig. 34 also shows that sera from many patients infected with HAV and HBV did not contain anti-NANB$_{5-1-1}$ antibodies, and that these antibodies were also not present in the sera from "normal" controls. Although one HAV patient (lane 36) appears to contain anti-NANB$_{5-1-1}$ antibodies, it is possible that this patient had been previously infected with HCV, since the incidence of NANBH is very high and since it is often subclinical.

These serological studies indicate that the cDNA in clone 5-1-1 encodes epitopes which are recognized specifically by sera from patients and animals infected with BB-NANBV. In addition, the cDNA does not appear to be derived from the primate genome. A hybridization probe made from clone 5-1-1 or from clone 81 did not hybridize to "Southern" blots of control human and chimpanzee genomic DNA from uninfected individuals under conditions where unique, single-copy genes are detectable. These probes also did not hybridize to Southern blots of control bovine genomic DNA.

IV.B.4. <u>Expression of the Polypeptide Encoded in a Composite of the HCV cDNAs in Clones 36, 81 and 32</u>

The HCV polypeptide which is encoded in the ORF which extends through clones 36, 81 and 32 was expressed as a fusion polypeptide with SOD. This was accomplished by inserting the composite cDNA, C100, into an expression cassette which contains the human superoxide dismutase gene, inserting the expression cassette into a yeast

expression vector, and expressing the polypeptide in yeast.

An expression cassette containing the composite C100 cDNA derived from clones 36, 81, and 32, was constructed by inserting the ~1270bp EcoRI fragment into the EcoRI site of the vector pS3-56 (also called pS356), yielding the plasmid pS3-56$_{C100}$. The construction of C100 is described in Section IV.A.16, supra.

The vector pS3-56, which is a pBR322 derivative, contains an expression cassette which is comprised of the ADH2/GAPDH hybrid yeast promoter upstream of the human superoxide dismutase gene, and a downstream GAPDH transcription terminator. A similar cassette, which contains these control elements and the superoxide dismutase gene has been described in Cousens et al. (1987), and in copending application EPO 196,056, published October 1, 1986, which is commonly owned by the herein assignee. The cassette in pS3-56, however, differs from that in Cousens et al. (1987) in that the heterologous proinsulin gene and the immunoglobulin hinge are deleted, and in that the gln$_{154}$ of the superoxide dismutase is followed by an adaptor sequence which contains an EcoRI site. The sequence of the adaptor is:

```
5'-AAT TTG GGA ATT CCA TAA TGA G      -3'
            AC CCT TAA GGT ATT ACT CAG CT    .
```

The EcoRI site allows the insertion of heterologous sequences which, when expressed from a vector containing the cassette, yield polypeptides which are fused to superoxide dismutase via an oligopeptide linker containing the amino acid sequence:

-asn-leu-gly-ile-arg-.

A sample of pS356 has been deposited on 29 April 1988 under the terms of the Budapest Treaty with the American Type Culture Collection (ATCC), 12301 Parklawn Dr., Rockville, Maryland 20853, and has been assigned Accession No. 67683. The terms and conditions for availability and access to the deposit, and for maintenance of the deposit are the same as those specified in Section II.A., for strains containing NANBV-cDNAs. This deposit is intended for convenience only, and is not required to practice the present invention in view of the description here. The deposited material is hereby incorporated herein by reference.

After recombinants containing the C100 cDNA insert in the correct orientation were isolated, the expression cassette containing the C100 cDNA was excised from pS3-56_{C100} with BamHI, and a fragment of ~3400bp which contains the cassette was isolated and purified. This fragment was then inserted into the BamHI site of the yeast vector pAB24.

Plasmid pAB24, the significant features of which are shown in Fig. 35, is a yeast shuttle vector which contains the complete 2 micron sequence for replication [Broach (1981)] and pBR322 sequences. It also contains the yeast URA3 gene derived from plasmid YEp24 [Botstein et al. (1979)], and the yeast LEU^{2d} gene derived from plasmid pC1/1. EPO Pub. No. 116,201. Plasmid pAB24 was constructed by digesting YEp24 with EcoRI and religating the vector to remove the partial 2 micron sequences. The resulting plasmid, YEP24deltaRI, was linearized by digestion with ClaI and ligated with the complete 2 micron plasmid which had been linearized with ClaI. The resulting plasmid, pCBou, was then digested with XbaI and the 8605 bp vector fragment was gel isolated. This isolated XbaI fragment was ligated with a 4460 bp XbaI fragment containing the LEU^{2d} gene isolated from pC1/1;

the orientation of the LEU^{2d} gene is in the same direction as the URA3 gene. Insertion of the expression was in the unique BamHI site of the pBR322 sequence, thus interrupting the gene for bacterial resistance to tetracycline.

The recombinant plasmid which contained the SOD-C100 expression cassette, pAB24C100-3, was transformed into yeast strain JSC 308, as well as into other yeast strains. The cells were transformed as described by Hinnen et al. (1978), and plated onto ura-selective plates. Single colonies were inoculated into leu-selective media and grown to saturation. The culture was induced to express the SOD-C100 polypeptide (called C100-3) by growth in YEP containing 1% glucose.

Strain JSC 308 is of the genotype MAT @, leu2, ura3(del) DM15 (GAP/ADR1) integrated at the ADR1 locus. In JSC 308, over-expression of the positive activator gene product, ADR1, results in hyperderepression (relative to an ADR1 wild type control) and significantly higher yields of expressed heterologous proteins when such proteins are synthesized via an ADH2 UAS regulatory system. The construction of the yeast strain JSC 308 is disclosed in copending application, U.S. Serial No. (Attorney Docket No. 2300-0229), filed concurrently herewith, and which is hereby incorporated herein by reference. A sample of JSC 308 has been deposited on 5 May 1988 with the ATCC under the conditions of the Budapest Treaty, and has been assigned Accession No. 20879. The terms and conditions for availability and access to the deposit, and for maintenance of the deposit are the same as those specified in Section II.A., for strains containing HCV cDNAs.

The complete C100-3 fusion polypeptide encoded in pAB24C100-3 should contain 154 amino acids of human SOD at the amino-terminus, 5 amino acid residues derived from the synthetic adaptor containing the EcoRI site, 363 amino

acid residues derived from C100 cDNA, and 5 carboxy-terminal amino acids derived from the MS2 nucleotide sequence adjoining the HCV cDNA sequence in clone 32. (See Section IV.A.7.) The putative amino acid sequence of the carboxy-terminus of this polypeptide, beginning at the penultimate Ala residue of SOD, is shown in Fig. 36; also shown is the nucleotide sequence encoding this portion of the polypeptide.

IV.B.5. Identification of the Polypeptide Encoded within C100 as an NANBH Associated Antigen

The C100-3 fusion polypeptide expressed from plasmid pAB24C100-3 in yeast strain JSC 308 was characterized with respect to size, and the polypeptide encoded within C100 was identified as an NANBH-associated antigen by its immunological reactivity with serum from a human with chronic NANBH.

The C100-3 polypeptide, which was expressed as described in Section IV.B.4., was analyzed as follows. Yeast JSC 308 cells were transformed with pAB24, or with pAB24C100-3, and were induced to express the heterologous plasmid encoded polypeptide. The induced yeast cells in 1 ml of culture ($OD_{650\ nm}$ ~20) were pelleted by centrifugation at 10,000 rpm for 1 minute, and were lysed by vortexing them vigorously (10 x 1 min) with 2 volumes of solution and 1 volume of glass beads (0.2 millimicron diameter). The solution contained 50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1mM phenylmethylsulphonyl fluoride (PMSF), and 1 microgram/ml pepstatin. Insoluble material in the lysate, which includes the C100-3 polypeptide, was collected by centrifugation (10,000 rpm for 5 minutes), and was dissolved by boiling for 5 minutes in Laemmli SDS sample buffer. [See Laemmli (1970)]. An amount of polypeptides equivalent to that in 0.3 ml of the induced yeast culture was subjected to electrophoresis through 10%

polyacrylamide gels in the presence of SDS according to Laemmli (1970). Protein standards were co-electrophoresed on the gels. Gels containing the expressed polypeptides were either stained with Coomassie brilliant blue, or were subjected to "Western" blotting as described in Section IV.B.2., using serum from a patient with chronic NANBH to determine the immunological reactivity of the polypeptides expressed from pAB24 and from pAB24C100-3.

The results are shown in Fig. 37. In Fig. 37A the polypeptides were stained with Coomassie brilliant blue. The insoluble polypeptide(s) from JSC 308 transformed with pAB24 and from two different colonies of JSC transformed with pAB24C100-3 are shown in lane 1 (pAB24), and lanes 2 and 3, respectively. A comparison of lanes 2 and 3 with lane 1 shows the induced expression of a polypeptide corresponding to a molecular weight of ~54,000 daltons from JSC 308 transformed with pAB24C100-3, which is not induced in JSC 308 transformed with pAB24. This polypeptide is indicated by the arrow.

Fig. 37B shows the results of the Western blots of the insoluble polypeptides expressed in JSC 308 transformed with pAB24 (lane 1), or with pAB24C100-3 (lane 2). The polypeptides expressed from pAB24 were not immunologically reactive with serum from a human with NANBH. However, as indicated by the arrow, JSC 308 transformed with pAB24C100-3 expressed a polypeptide of ~54,000 dalton molecular weight which did react immunologically with the human NANBH serum. The other immunologically reactive polypeptides in lane 2 may be degradation and/or aggregation products of this ~54,000 dalton polypeptide.

IV.B.6. <u>Purification of Fusion Polypeptide C100-3</u>

The fusion polypeptide, C100-3, comprised of SOD at the N-terminus and in-frame C100 HCV-polypeptide at the C-terminus was purified by differential extraction of the

insoluble fraction of the extracted host yeast cells in which the polypeptide was expressed.

The fusion polypeptide, C100-3, was expressed in yeast strain JSC 308 transformed with pAB24C100-3, as described in Section IV.B.4. The yeast cells were then lysed by homogenization, the insoluble material in the lysate was extracted at pH 12.0, and C100-3 in the remaining insoluble fraction was solubilized in buffer containing SDS.

The yeast lysate was prepared essentially according to Nagahuma et al. (1984). A yeast cell suspension was prepared which was 33% cells (v/v) suspended in a solution (Buffer A) containing 20 mM Tris HCl, pH 8.0, 1 mM dithiothreitol, and 1 mM phenylmethylsulfonylfluoride (PMSF). An aliquot of the suspension (15 ml) was mixed with an equal volume of glass beads (0.45-0.50 mm diameter), and the mixture was vortexed at top speed on a Super Mixer (Lab Line Instruments, Inc.) for 8 min. The homogenate and glass beads were separated, and the glass beads were washed 3 times with the same volume of Buffer A as the original packed cells. After combining the washes and homogenate, the insoluble material in the lysate was obtained by centrifuging the homogenate at 7,000 x g for 15 minutes at 4^{o}C, resuspending the pellets in Buffer A equal to twice the volume of original packed cells, and re-pelleting the material by centrifugation at 7,000 x g for 15 min. This washing procedure was repeated 3 times.

The insoluble material from the lysate was extracted at pH 12.0 as follows. The pellet was suspended in buffer containing 0.5 M NaCl, 1 mM EDTA, where the suspending volume was equal to 1.8 times the of the original packed cells. The pH of the suspension was adjusted by adding 0.2 volumes of 0.4 M Na phosphate buffer, pH 12.0. After mixing, the suspension was centrifuged at 7,000 x g for 15 min at 4^{o}C, and the super-

natant removed. The extraction was repeated 2 times. The extracted pellets were washed by suspending them in 0.5 M NaCl, 1 mM EDTA, using a suspension volume equal to two volumes of the original packed cells, followed by centrifugation at 7,000 x g for 15 min at 4°C.

The C100-3 polypeptide in the extracted pellet was solubilized by treatment with SDS. The pellets were suspended in Buffer A equal to 0.9 volumes of the original packed cell volume, and 0.1 volumes of 2% SDS was added. After the suspension was mixed, it was centrifuged at 7,000 x g for 15 min at 4°C. The resulting pellet was extracted 3 more times with SDS. The resulting supernatants, which contained C100-3 were pooled.

This procedure purifies C100-3 more than 10-fold from the insoluble fraction of the yeast homogenate, and the recovery of the polypeptide is greater than 50%.,

The purified preparation of fusion polypeptide was analyzed by polyacrylamide gel electrophoresis according to Laemmli (1970). Based upon this analysis, the polypeptide was greater than 80% pure, and had an apparent molecular weight of ~54,000 daltons.

IV.C. <u>Identification of RNA in Infected Individuals Which Hybridizes to HCV cDNA</u>.

IV.C.1. <u>Identification of RNA in the Liver of a Chimpanzee With NANBH Which Hybridizes to HCV cDNA</u>.

RNA from the liver of a chimpanzee which had NANBH was shown to contain a species of RNA which hybridized to the HCV cDNA contained within clone 81 by Northern blotting, as follows.

RNA was isolated from a liver biopsy of the chimpanzee from which the high titer plasma was derived (see Section IV.A.1.) using techniques described in Maniatis et al. (1982) for the isolation of total RNA from

mammalian cells, and for its separation into poly A^{+} and poly A^{-} fractions. These RNA fractions were subjected to electrophoresis on a formaldehyde/agarose gel (1% w/v), and transferred to nitrocellulose. (Maniatis et al. (1982)). The nitrocellulose filters were hybridized with radiolabeled HCV cDNA from clone 81 (see Fig. 4 for the nucleotide sequence of the insert.) To prepare the radiolabeled probe, the HCV cDNA insert isolated from clone 81 was radiolabeled with ^{32}P by nick translation using DNA Polymerase I (Maniatis et al. (1982)). Hybridization was for 18 hours at $42^{o}C$ in a solution containing 10% (w/v) Dextran sulphate, 50% (w/v) deionized formamide, 750 mM NaCl, 75 mM Na citrate, 20 mM Na_2HPO_4, pH 6.5, 0.1% SDS, 0.02% (w/v) bovine serum albumin (BSA), 0.02% (w/v) Ficoll-400, 0.02% (w/v) polyvinylpyrrolidone, 100 micrograms/ml salmon sperm DNA which had been sheared by sonication and denatured, and 10^{6} CPM/ml of the nick-translated cDNA probe.

An autoradiograph of the probed filter is shown in Fig. 38. Lane 1 contains ^{32}P-labeled restriction fragment markers. Lanes 2-4 contain chimpanzee liver RNA as follows: lane 2 contains 30 micrograms of total RNA; lane 3 contains 30 micrograms of poly A- RNA; and lane 4 contains 20 micrograms of poly A+ RNA. As shown in Fig. 38, the liver of the chimpanzee with NANBH contains a heterogeneous population of related poly A+ RNA molecules which hybridizes to the HCV cDNA probe, and which appears to be from about 5000 nucleotides to about 11,000 nucleotides in size. This RNA, which hybridizes to the HCV cDNA, could represent viral genomes and/or specific transcripts of the viral genome.

The experiment described in Section IV.C.2., infra, is consistent with the suggestion that HCV contains an RNA genome.

IV.C.2. Identification of HCV Derived RNA in Serum from Infected Individuals.

Nucleic acids were extracted from particles isolated from high titer chimpanzee NANBH plasma as described in Section IV.A.1.. Aliquots (equivalent to 1 ml of original plasma) of the isolated nucleic acids were resuspended in 20 microliters 50 mM Hepes, pH 7.5, 1 mm EDTA and 16 micrograms/ml yeast soluble RNA. The samples were denatured by boiling for 5 minutes followed by immediate freezing, and were treated with RNase A (5 microliters containing 0.1 mg/ml RNase A in 25 mM EDTA, 40 mM Hepes, pH 7.5) or with DNase I (5 microliters containing 1 unit DNase I in 10 mM $MgCl_2$, 25 mM Hepes, pH 7.5); control samples were incubated without enzyme. Following incubation, 230 microliters of ice-cold 2XSSC containing 2 micrograms/ml yeast soluble RNA was added, and the samples were filtered on a nitrocellulose filter. The filters were hybridized with a cDNA probe from clone 81, which had been ^{32}P-labeled by nick-translation. Fig. 39 shows an autoradiograph of the filter. Hybridization signals were detected in the DNase treated and control samples (lanes 2 and 1, respectively), but were not detected in the RNase treated sample (lane 3). Thus, since RNase A treatment destroyed the nucleic acids isolated from the particles, and DNase I treatment had no effect, the evidence strongly suggests that the HCV genome is composed of RNA.

IV.C.3. Detection of Amplified HCV Nucleic Acid Sequences derived from HCV Nucleic Acid Sequences in Liver and Plasma Specimens from Chimpanzees with NANBH

HCV nucleic acids present in liver and plasma of chimpanzees with NANBH, and in control chimpanzees, were amplified using essentially the polymerase chain reaction (PCR) technique described by Saiki et al. (1986). The primer oligonucleotides were derived from the HCV cDNA

sequences in clone 81, or clones 36 and 37. The amplified sequences were detected by gel electrophoresis and Southern blotting, using as probes the appropriate cDNA oligomer with a sequence from the region between, but not including, the two primers.

Samples of RNA containing HCV sequences to be examined by the amplification system were isolated from liver biopsies of three chimpanzees with NANBH, and from two control chimpanzees. The isolation of the RNA fraction was by the guanidinium thiocyanate procedure described in Section IV.C.1.

Samples of RNA which were to be examined by the amplification system were also isolated from the plasmas of two chimpanzees with NANBH, and from one control chimpanzee, as well as from a pool of plasmas from control chimpanzees. One infected chimpanzee had a CID/ml equal to or greater than 10^6, and the other infected chimpanzee had a CID/ml equal to or greater than 10^5.

The nucleic acids were extracted from the plasma as follows. Either 0.1 ml or 0.01 ml of plasma was diluted to a final volume of 1.0 ml, with a TENB/proteinase K/SDS solution (0.05 M Tris-HCL, pH 8.0, 0.001 M EDTA, 0.1 M NaCl, 1 mg/ml Proteinase K, and 0.5% SDS) containing 10 micrograms/ml polyadenylic acid, and incubated at 37°C for 60 minutes. After this proteinase K digestion, the resultant plasma fractions were deproteinized by extraction with TE (10.0 mM Tris-HCl, pH 8.0, 1 mM EDTA) saturated phenol. The phenol phase was separated by centrifugation, and was reextracted with TENB containing 0.1% SDS. The resulting aqueous phases from each extraction were pooled, and extracted twice with an equal volume of phenol/chloroform/isoamyl alcohol [1:1(99:2)], and then twice with an equal volume of a 99:1 mixture of chloroform/isoamyl alcohol. Following phase separation by centrifugation, the aqueous phase was

brought to a final concentration of 0.2 M Na Acetate, and the nucleic acids were precipitated by the addition of two volumes of ethanol. The precipitated nucleic acids were recovered by ultracentrifugation in a SW 41 rotor at 38 K, for 60 minutes at 4°C.

In addition to the above, the high titer chimpanzee plasma and the pooled control plasma alternatively were extracted with 50 micrograms of poly A carrier by the procedure of Chomcyzski and Sacchi (1987). This procedure uses an acid guanidinium thiocyanate extraction. RNA was recovered by centrifugation at 10,000 RPM for 10 minutes at 4°C in an Eppendorf microfuge.

On two occasions, prior to the synthesis of cDNA in the PCR reaction, the nucleic acids extracted from plasma by the proteinase K/SDS/phenol method were further purified by binding to and elution from S and S Elutip-R Columns. The procedure followed was according to the manufacturer's directions.

The cDNA used as a template for the PCR reaction was derived from the nucleic acids (either total nucleic acids or RNA) prepared as described above. Following ethanol precipitation, the precipitated nucleic acids were dried, and resuspended in DEPC treated distilled water. Secondary structures in the nucleic acids were disrupted by heating at 65°C for 10 minutes, and the samples were immediately cooled on ice. cDNA was synthesized using 1 to 3 micrograms of total chimpanzee RNA from liver, or from nucleic acids (or RNA) extracted from 10 to 100 microliters of plasma. The synthesis utilized reverse transcriptase, and was in a 25 microliter reaction, using the protocol specified by the manufacturer, BRL. The primers for cDNA synthesis were those also utilized in the PCR reaction, described below. All reaction mixtures for cDNA synthesis contained 23 units of the RNAase inhibitor, RNASIN™ (Fisher/Promega). Following cDNA synthesis, the

reaction mixtures were diluted with water, boiled for 10 minutes, and quickly chilled on ice.

The PCR reactions were performed essentially according to the manufacturer's directions (Cetus-Perkin-Elmer), except for the addition of 1 microgram of RNase A. The reactions were carried out in a final volume of 100 microliters. The PCR was performed for 35 cycles, utilizing a regimen of 37°C, 72°C, and 94°C.

The primers for cDNA synthesis and for the PCR reactions were derived from the HCV cDNA sequences in either clone 81, clone 36, or clone 37b. (The HCV cDNA sequences of clones 81, 36, and 37b are shown in Figs. 4, 5, and 10, respectively.) The sequences of the two 16-mer primers derived from clone 81 were:

5' CAA TCA TAC CTG ACA G 3'

and

5' GAT AAC CTC TGC CTG A 3'.

The sequence of the primer from clone 36 was:

5' GCA TGT CAT GAT GTA T 3'.

The sequence of the primer from clone 37b was:

5' ACA ATA CGT GTG TCA C 3'.

In the PCR reactions, the primer pairs consisted of either the two 16-mers derived from clone 81, or the 16-mer from clone 36 and the 16-mer from clone 37b.

The PCR reaction products were analyzed by separation of the products by alkaline gel electrophoresis, followed by Southern blotting, and detection of the amplified HCV-cDNA sequences with a ^{32}P-labeled internal oligonucleotide probe derived from a

region of the HCV cDNA which does not overlap the primers. The PCR reaction mixtures were extracted with phenol/ chloroform, and the nucleic acids precipitated from the aqueous phase with salt and ethanol. The precipitated nucleic acids were collected by centrifugation, and dissolved in distilled water. Aliquots of the samples were subjected to electrophoresis on 1.8% alkaline agarose gels. Single stranded DNA of 60, 108, and 161 nucleotide lengths were co-electrophoresed on the gels as molecular weight markers. After electrophoresis, the DNAs in the gel were transferred onto Biorad Zeta Probe™ paper. Prehybridization and hybridization, and wash conditions were those specified by the manufacturer (Biorad).

The probes used for the hybridization-detection of amplified HCV cDNA sequences were the following. When the pair of PCR primers were derived from clone 81, the probe was an 108-mer with a sequence corresponding to that which is located in the region between the sequences of the two primers. When the pair of PCR primers were derived from clones 36 and 37b, the probe was the nick-translated HCV cDNA insert derived from clone 35. The primers are derived from nucleotides 155-170 of the clone 37b insert, and 206-268 of the clone 36 insert. The 3'-end of the HCV cDNA insert in clone 35 overlaps nucleotides 1-186 of the insert in clone 36; and the 5'-end of clone 35 insert overlaps nucleotides 207-269 of the insert in clone 37b. (Compare Figs. 5, 8 and 10.) Thus, the cDNA insert in clone 35 spans part of the region between the sequences of the clone 36 and 37b derived primers, and is useful as a probe for the amplified sequences which include these primers.

Analysis of the RNA from the liver specimens was according to the above procedure utilizing both sets of primers and probes. The RNA from the liver of the three chimpanzees with NANBH yielded positive hybridization

results for amplification sequences of the expected size (161 and 586 nucleotides for 81 and 36 and 37b, respectively), while the control chimpanzees yielded negative hybridization results. The same results were achieved when the experiment was repeated three times.

Analysis of the nucleic acids and RNA from plasma was also according to the above procedure utilizing the primers and probe from clone 81. The plasmas were from two chimpanzees with NANBH, from a control chimpanzee, and pooled plasmas from control chimpanzees. Both of the NANBH plasmas contained nucleic acids/RNA which yielded positive results in the PCR amplified assay, while both of the control plasmas yielded negative results. These results have been repeatably obtained several times.

IV.D. Radioimmunoassay for Detecting HCV Antibodies in Serum from Infected Individuals

Solid phase radioimmunoassays to detect antibodies to HCV antigens were developed based upon Tsu and Herzenberg (1980). Microtiter plates (Immulon 2, Removawell strips) are coated with purified polypeptides containing HCV epitopes. The coated plates are incubated with either human serum samples suspected of containing antibodies to the HCV epitopes, or to appropriate controls. During incubation, antibody, if present, is immunologically bound to the solid phase antigen. After removal of the unbound material and washing of the microtiter plates, complexes of human antibody-NANBV antigen are detected by incubation with ^{125}I-labeled sheep anti-human immunoglobulin. Unbound labeled antibody is removed by aspiration, and the plates are washed. The radioactivity in individual wells is determined; the amount of bound human anti-HCV antibody is proportional to the radioactivity in the well.

IV.D.1. <u>Purification of Fusion Polypeptide SOD-NANB$_{5-1-1}$</u>.

The fusion polypeptide SOD-NANB$_{5-1-1}$, expressed in recombinant bacteria as described in Section IV.B.1., was purified from the recombinant *E. coli* by differential extraction of the cell extracts with urea, followed by chromatography on anion and cation exchange columns as follows.

Thawed cells from 1 liter of culture were resuspended in 10 ml of 20% (w/v) sucrose containing 0.01M Tris HCl, pH 8.0, and 0.4 ml of 0.5M EDTA, pH 8.0 was added. After 5 minutes at 0°C, the mixture was centrifuged at 4,000 x g for 10 minutes. The resulting pellet was suspended in 10 ml of 25% (w/v) sucrose containing 0.05 M Tris HCl, pH 8.0, 1 mM phenylmethylsulfonylfluoride (PMSF) and 1 microgram/ml pepstatin A, followed by addition of 0.5 ml lysozyme (10 mg/ml) and incubation at 0°C for 10 minutes. After the addition of 10 ml 1% (v/v) Triton X-100 in 0.05 M Tris HCl, pH 8.0, 1 mM EDTA, the mixture was incubated an additional 10 min at 0°C with occasional shaking. The resulting viscous solution was homogenized by passage 6 times through a sterile 20-gauge hypodermic needle, and centrifuged at 13,000 x g for 25 minutes. The pelleted material was suspended in 5 ml of 0.01 M Tris HCl pH 8.0, and the suspension centrifuged at 4,000 x g for 10 minutes. The pellet, which contained SOD-NANB$_{5-1-1}$ fusion protein, was dissolved in 5 ml of 6 M urea in 0.02 M Tris HCl, pH 8.0, 1 mM dithiothreitol (Buffer A), and was applied to a column of Q-Sepharose Fast Flow equilibrated with Buffer A. Polypeptides were eluted with a linear gradient of 0.0 to 0.3 M NaCl in Buffer A. After elution, fractions were analyzed by polyacrylamide gel electrophoresis in the presence of SDS to determine their content of SOD-NANB$_{5-1-1}$. Fractions containing this polypeptide were pooled, and dialyzed against 6 M urea in

0.02 M sodium phosphate buffer, pH 6.0, 1 mM dithiothreitol (Buffer B). The dialyzed sample was applied on a column of S-Sepharose Fast Flow equilibrated with Buffer B, and polypeptides eluted with a linear gradient of 0.0 to 0.3 M NaCl in Buffer B. The fractions were analyzed by polyacrylamide gel electrophoresis for the presence of SOD-NANB$_{5-1-1}$, and the appropriate fractions were pooled.

The final preparation of SOD-NANB$_{5-1-1}$ polypeptide was examined by electrophoresis on polyacrylamide gels in the presence of SDS. Based upon this analysis, the preparation was more than 80% pure.

IV.D.2. Purification of Fusion Polypeptide SOD-NANB$_{81}$.

The fusion polypeptide SOD-NANB$_{81}$, expressed in recombinant bacteria as described in Section IV.B.2., was purified from recombinant *E. coli* by differential extraction of the cell extracts with urea, followed by chromatography on anion and cation exchange columns utilizing the procedure described for the isolation of fusion polypeptide SOD-NANB$_{5-1-1}$ (See Section IV.D.1.).

The final preparation of SOD-NANB$_{81}$ polypeptide was examined by electrophoresis on polyacrylamide gels in the presence of SDS. Based upon this analysis, the preparation was more than 50% pure.

IV.D.3. Detection of Antibodies to HCV Epitopes by Solid Phase Radioimmunoassay.

Serum samples from 32 patients who were diagnosed as having NANBH were analyzed by radioimmunoassay (RIA) to determine whether antibodies to HCV epitopes present in fusion polypeptides SOD-NANB$_{5-1-1}$ and SOD-NANB$_{81}$ were detected.

Microtiter plates were coated with SOD-NANB$_{5-1-1}$ or SOD-NANB$_{81}$, which had been partially purified according

to Sections IV.D.1. and IV.D.2., respectively. The assays were conducted as follows.

One hundred microliter aliquots containing 0.1 to 0.5 micrograms of SOD-$NANB_{5-1-1}$ or SOD-$NANB_{81}$ in 0.125 M Na borate buffer, pH 8.3, 0.075 M NaCl (BBS) was added to each well of a microtiter plate (Dynatech Immulon 2 Removawell Strips). The plate was incubated at 4°C overnight in a humid chamber, after which, the protein solution was removed and the wells washed 3 times with BBS containing 0.02% Triton X-100 (BBST). To prevent non-specific binding, the wells were coated with bovine serum albumin (BSA) by addition of 100 microliters of a 5 mg/ml solution of BSA in BBS followed by incubation at room temperature for 1 hour; after this incubation the BSA solution was removed. The polypeptides in the coated wells were reacted with serum by adding 100 microliters of serum samples diluted 1:100 in 0.01M Na phosphate buffer, pH 7.2, 0.15 M NaCl (PBS) containing 10 mg/ml BSA, and incubating the serum containing wells for 1 hr at 37°C. After incubation, the serum samples were removed by aspiration, and the wells were washed 5 times with BBST. Anti-$NANB_{5-1-1}$ and Anti-$NANB_{81}$ bound to the fusion polypeptides was determined by the binding of ^{125}I-labeled $F'(ab)_2$ sheep anti-human IgG to the coated wells. Aliquots of 100 microliters of the labeled probe (specific activity 5-20 microcuries/microgram) were added to each well, and the plates were incubated at 37°C for 1 hour, followed by removal of excess probe by aspiration, and 5 washes with BBST. The amount of radioactivity bound in each well was determined by counting in a counter which detects gamma radiation.

The results of the detection of anti-$NANB_{5-1-1}$ and anti-$NANB_{81}$ in individuals with NANBH is presented in Table 1.

Table 1
Detection of Anti-5-1-1 and Anti-81 in Sera of NANB, HAV and HBV Hepatitis Patients

Patient Reference Number	Diagnosis	S/N Anti-5-1-1	S/N Anti-81
1. 28[1]	Chronic NANB, IVD[2]	0.77	4.20
	Chronic NANB, IVD	1.14	5.14
	Chronic NANB, IVD	2.11	4.05
2. 29[1]	AVH[3], NANB, Sporadic	1.09	1.05
	Chronic, NANB	33.89	11.39
	Chronic, NANB	36.22	13.67
3. 30[1]	AVH, NANB, IVD	1.90	1.54
	Chronic NANB, IVD	34.17	30.28
	Chronic NANB, IVD	32.45	30.84
4. 31	Chronic NANB, PT[4]	16.09	8.05
5. 32[1]	Late AVH NANB, IVD	0.69	0.94
	Late AVH NANB, IVD	0.73	0.68
6. 33[1]	AVH, NANB, IVD	1.66	1.96
	AVH, NANB, IVD	1.53	0.56
7. 34[1]	Chronic NANB, PT	34.40	7.55
	Chronic NANB, PT	45.55	13.11
	Chronic NANB, PT	41.58	13.45
	Chronic NANB, PT	44.20	15.48
8. 35[1]	AVH NANB, IVD	31.92	31.95
	"Healed" recent NANB, AVH	6.87	4.45
9. 36	Late AVH NANB PT	11.84	5.79
10. 37	AVH NANB, IVD	6.52	1.33
11. 38	Late AVH NANB, PT	39.44	39.18
12. 39	Chronic NANB, PT	42.22	37.54
13. 40	AVH, NANB, PT	1.35	1.17
14. 41	Chronic NANB? PT	0.35	0.28

Patient Reference Number	Diagnosis	S/N Anti-5-1-1	Anti-81
15. 42	AVH, NANB, IVD	6.25	2.34
16. 43	Chronic NANB, PT	0.74	0.61
17. 44	AVH, NANB, PT	5.40	1.83
18. 45	Chronic, NANB, PT	0.52	0.32
19. 46	AVH, NANB	23.35	4.45
20. 47	AVH, Type A	1.60	1.35
21. 48	AVH, Type A	1.30	0.66
22. 49	AVH, Type A	1.44	0.74
23. 50	Resolved Recent AVH, Type A	0.48	0.56
24. 51	AVH, Type A	0.68	0.64
	Resolved AVH, Type A	0.80	0.65
25. 52	Resolved Recent AVH, Type A	1.38	1.04
	Resolved Recent AVH, Type A	0.80	0.65
26. 53	AVH, Type A	1.85	1.16
	Resolved Recent AVH, Type A	1.02	0.88
27. 54	AVH, Type A	1.35	0.74
28. 55	Late AVH, HBV	0.58	0.55
29. 56	Chronic HBV	0.84	1.06
30. 57	Late AVH, HBV	3.20	1.60
31. 58	Chronic HBV	0.47	0.46
32. 59[1]	AVH, HBV	0.73	0.60
	Healed AVH, HBV	0.43	0.44
33. 60[1]	AVH, HBV	1.06	0.92
	Healed AVH, HBV	0.75	0.68

Patient Reference Number	Diagnosis	S/N Anti-5-1-1	Anti-81
34. 61[1]	AVH, HBV	1.66	0.61
	Healed AVH, HBV	0.63	0.36
35. 62[1]	AVH, HBV	1.02	0.73
	Healed AVH, HBV	0.41	0.42
36. 63[1]	AVH, HBV	1.24	1.31
	Healed AVH, HBV	1.55	0.45
37. 64[1]	AVH, HBV	0.82	0.79
	Healed AVH, HBV	0.53	0.37
38. 65[1]	AVH, HBV	0.95	0.92
	Healed AVH, HBV	0.70	0.50
39. 66[1]	AVH, HBV	1.03	0.68
	Healed AVH, HBV	1.71	1.39

[1]Sequential serum samples available from these patients
[2]IVD=Intravenus Drug User
[3]AVH=Acute viral hepatitis
[4]PT=Post transfusion

As seen in Table 1, 19 of 32 sera from patients diagnosed as having NANBH were positive with respect to antibodies directed against HCV epitopes present in SOD-$NANB_{5-1-1}$ and SOD-$NANB_{81}$.

However, the serum samples which were positive were not equally immunologically reactive with SOD-$NANB_{5-1-1}$ and SOD-$NANB_{81}$. Serum samples from patient No. 1 were positive to SOD-$NANB_{81}$ but not to SOD-$NANB_{5-1-1}$. Serum samples from patients number 10, 15, and 17 were positive to SOD-$NANB_{5-1-1}$ but not to SOD-$NANB_{81}$. Serum samples from patients No. 3, 8, 11, and 12 reacted equally with both fusion polypeptides, whereas serum samples from patients No. 2, 4, 7, and 9 were 2-3 fold higher in the reaction to SOD-$NANB_{5-1-1}$ than to SOD-$NANB_{81}$. These results suggest that $NANB_{5-1-1}$ and $NANB_{81}$ may contain at least 3 different epitopes; i.e., it is possible that each polypeptide contains at least 1 unique epitope, and that the two polypeptides share at least 1 epitope.

IV.D.4. <u>Specificity of the Solid Phase RIA for NANBH</u>

The specificity of the solid phase RIAs for NANBH was tested by using the assay on serum from patients infected with HAV or with HBV and on sera from control individuals. The assays utilizing partially purified SOD-$NANB_{5-1-1}$ and SOD-$NANB_{81}$ were conducted essentially as described in Section IV.D.3, except that the sera was from patients previously diagnosed as having HAV or HBV, or from individuals who were blood bank donors. The results for sera from HAV and HBV infected patients are presented in table 1. The RIA was tested using 11 serum specimens from HAV infected patients, and 20 serum specimens from HBV infected patients. As shown in table 1, none of these sera yielded a positive immunological reaction with the fusion polypeptides containing BB-NANBV epitopes.

The RIA using the $NANB_{5-1-1}$ antigen was used to determine immunological reactivity of serum from control individuals. Out of 230 serum samples obtained from the normal blood donor population, only 2 yielded positive reactions in the RIA (data not shown). It is possible that the two blood donors from whom these serum samples originated had previously been exposed to HCV.

IV.D.5. Reactivity of $NANB_{5-1-1}$ During the Course of NANBH Infection.

The presence of anti-$NANB_{5-1-1}$ antibodies during the course of NANBH infection of 2 patients and 4 chimpanzees was followed using RIA as described in Section IV.D.3. In addition the RIA was used to determine the presence or absence of anti-$NANB_{5-1-1}$ antibodies during the course of infection of HAV and HBV in infected chimpanzees.

The results, which are presented in Table 2, show that with chimpanzees and with humans, anti-$NANB_{5-1-1}$ antibodies were detected following the onset of the acute phase of NANBH infection. Anti-$NANB_{5-1-1}$ antibodies were not detected in serum samples from chimpanzees infected with either HAV or HBV. Thus anti-$NANB_{5-1-1}$ antibodies serve as a marker for an individual's exposure to HCV.

Table 2
Seroconversion in Sequential Serum Samples from Hepatitis Patients and Chimpanzees Using 5-1-1 Antigen

Patient/ Chimp	Sample Date (Days) (o=inoculation day)	Hepatitis Viruses	Anti-5-1-1 (S/N)	ALT (mu/ml
Patient 29	T#	NANB	1.09	1180
	T+180		33.89	425
	T+208		36.22	--
Patient 30	T	NANB	1.90	1830
	T+307		34.17	290
	T+799		32.45	276
Chimp 1	0	NANB	0.87	9
	76		0.93	71
	118		23.67	19
	154		32.41	--
Chimp 2	0	NANB	1.00	5
	21		1.08	52
	73		4.64	13
	138		25.01	--
Chimp 3	0	NANB	1.08	8
	43		1.44	205
	53		1.82	14
	159		11.87	6
Chimp 4	-3	NANB	1.12	11
	55		1.25	132
	83		6.60	--
	140		17.51	--
Chimp 5	0	HAV	1.50	4
	25		2.39	147
	40		1.92	18
	268		1.53	5
Chimp 6	-8	HAV	0.85	--
	15		--	106
	41		0.81	10
	129		1.33	--

Patient/ Chimp	Sample Date (Days) (o=inoculation day)	Hepatitis Viruses	Anti-5-1-1 (S/N)	ALT (mu/ml)
Chimp 7	0	HAV	1.17	7
	22		1.60	83
	115		1.55	5
	139		1.60	--
Chimp 8	0	HAV	0.77	15
	26		1.98	130
	74		1.77	8
	205		1.27	5
Chimp 9	-290	HBV	1.74	--
	379		3.29	9
	435		2.77	6
Chimp 10	0	HBV	2.35	8
	111-118 (pool)		2.74	96-156 (pool)
	205		2.05	9
	240		1.78	13
Chimp 11	0	HBV	1.82	11
	28-56 (pool)		1.26	8-100 (pool)
	169		--	9
	223		0.52	10

*T=day of initial sampling

IV.E. Purification of Polyclonal Serum Antibodies to $NANB_{5-1-1}$

On the basis of the specific immunological reactivity of the SOD-$NANB_{5-1-1}$ polypeptide with the antibodies in serum samples from patients with NANBH, a method was developed to purify serum antibodies which react immunologically with the epitope(s) in $NANB_{5-1-1}$. This method utilizes affinity chromatography. Purified SOD-$NANB_{5-1-1}$ polypeptide (see Section IV.D.1) was attached to an insoluble support; the attachment is such that the immobilized polypeptide retains its affinity for antibody to $NANB_{5-1-1}$. Antibody in serum samples is absorbed to the matrix-bound polypeptide. After washing to remove non-specifically bound materials and unbound materials, the bound antibody is released from the bound SOD-HCV polypeptide by change in pH, and/or by chaotropic reagents, for example, urea.

Nitrocellulose membranes containing bound SOD-$NANB_{5-1-1}$ were prepared as follows. A nitrocellulose membrane, 2.1 cm Sartorius of 0.2 micron pore size, was washed for 3 minutes three times with BBS. SOD-$NANB_{5-1-1}$ was bound to the membrane by incubation of the purified preparation in BBS at room temperature for 2 hours; alternatively it was incubated at 4°C overnight. The solution containing unbound antigen was removed, and the filter was washed three times with BBS for three minutes per wash. The remaining active sites on the membrane were blocked with BSA by incubation with a 5 mg/ml BSA solution for 30 minutes. Excess BSA was removed by washing the membrane with 5 times with BBS and 3 times with distilled water. The membrane containing the viral antigen and BSA was then treated with 0.05 M glycine hydrochloride, pH 2.5, 0.10 M NaCl (GlyHCl) for 15 minutes, followed by 3 three minute washes with PBS.

Polyclonal anti-$NANB_{5-1-1}$ antibodies were isolated by incubating the membranes containing the fusion polypeptide with serum from an individual with NANBH for 2 hours. After the incubation, the filters were washed 5 times with BBS, and twice with distilled water. Bound antibodies were then eluted from each filter with 5 elutions of GlyHCl, at 3 minutes per elution. The pH of the eluates was adjusted to pH 8.0 by collecting each eluate in a test tube containing 2.0 M Tris HCl, pH 8.0. Recovery of the anti-$NANB_{5-1-1}$ antibody after affinity chromatography is approximately 50%.

The nitrocellulose membranes containing the bound viral antigen can be used several times without appreciable decrease in binding capacity. To reuse the membranes, after the antibodies have been eluted the membranes are washed with BBS three times for 3 minutes. They are then stored in BBS at $4^{O}C$.

IV.F. The Capture of HCV Particles from Infected Plasma Using Purified Human Polyclonal Anti-HCV Antibodies; Hybridization of the Nucleic Acid in the Captured Particles to HCV cDNA

IV.F.1. The Capture of HCV Particles from Infected Plasma Using Human Polyclonal Anti-HCV Antibodies

Protein-nucleic acid complexes present in infectious plasma of a chimpanzee with NANBH were isolated using purified human polyclonal anti-HCV antibodies which were bound to polystyrene beads.

Polyclonal anti-$NANB_{5-1-1}$ antibodies were purified from serum from a human with NANBH using the SOD-HCV polypeptide encoded in clone 5-1-1. The method for purification was that described in Section IV.E.

The purified anti-$NANB_{5-1-1}$ antibodies were bound to polystyrene beads (1/4" diameter, specular fin-

ish, Precision Plastic Ball Co., Chicago, Illinois) by incubating each at room temperature overnight with 1 ml of antibodies (1 microgram/ml in borate buffered saline, pH 8.5). Following the overnight incubation, the beads were washed once with TBST [50 mM Tris HCl, pH 8.0, 150 mM NaCl, 0.05% (v/v) Tween 20], and then with phosphate buffered saline (PBS) containing 10 mg/ml BSA.

Control beads were prepared in an identical fashion, except that the purified anti-$NANB_{5-1-1}$ antibodies were replaced with total human immunoglobulin.

Capture of HCV from NANBH infected chimpanzee plasma using the anti-$NANB_{5-1-1}$ antibodies bound to beads was accomplished as follows. The plasma from a chimpanzee with NANBH used is described in Section IV.A.1.. An aliquot (1 ml) of the NANBV infected chimpanzee plasma was incubated for 3 hours at 37°C with each of 5 beads coated with either anti-$NANB_{5-1-1}$ antibodies, or with control immunoglobulins. The beads were washed 3 times with TBST.

IV.F.2. Hybridization of the Nucleic Acid in the Captured Particles to NANBV-cDNA

The nucleic acid component released from the particles captured with anti-$NANB_{5-1-1}$ antibodies was analyzed for hybridization to HCV cDNA derived from clone 81.

HCV particles were captured from NANBH infected chimpanzee plasma, as described in IV.F.1. To release the nucleic acids from the particles, the washed beads were incubated for 60 min. at 37°C with 0.2 ml per bead of a solution containing proteinase k (1 mg/ml), 10 mM Tris HCl, pH 7.5, 10 mM EDTA, 0.25% (w/v) SDS, 10 micrograms/ml soluble yeast RNA, and the supernatant solution was removed. The supernatant was extracted with phenol and chloroform, and the nucleic acids precipitated with ethanol overnight at -20°C. The nucleic acid precipitate

was collected by centrifugation, dried, and dissolved in 50 mM Hepes, pH 7.5. Duplicate aliquots of the soluble nucleic acids from the samples obtained from beads coated with anti-$NANB_{5-1-1}$ antibodies and with control beads containing total human immunoglobulin were filtered onto to nitrocellulose filters. The filters were hybridized with a ^{32}P-labeled, nick-translated probe made from the purified HCV cDNA fragment in clone 81. The methods for preparing the probe and for the hybridization are described in Section IV.C.1..

Autoradiographs of a probed filter containing the nucleic acids from particles captured by beads containing anti-$NANB_{5-1-1}$ antibodies are shown in Fig. 40. The extract obtained using the anti-$NANB_{5-1-1}$ antibody (A_1,A_2) gave clear hybridization signals relative to the control antibody extract (A_3,A_4) and to control yeast RNA (B_1,B_2). Standards consisting of 1pg, 5pg, and 10pg of the purified, clone 81 cDNA fragment are shown in C1-3, respectively.

These results demonstrate that the particles captured from NANBH plasma by anti-$NANB_{5-1-1}$-antibodies contain nucleic acids which hybridize with HCV cDNA in clone 81, and thus provide further evidence that the cDNAs in these clones are derived from the etiologic agent for NANBH.

IV.G. Immunological Reactivity of C100-3 with Purified Anti-$NANB_{5-1-1}$ Antibodies

The immunological reactivity of C100-3 fusion polypeptide with anti-$NANB_{5-1-1}$ antibodies was determined by a radioimmunoassay, in which the antigens which were bound to a solid phase were challenged with purified anti-$NANB_{5-1-1}$ antibodies, and the antigen-antibody complex detected with ^{125}I-labeled sheep anti-human antibodies.

The immunological reactivity of C100-3 polypeptide was compared with that of SOD-NANB$_{5-1-1}$ antigen.

The fusion polypeptide C100-3 was synthesized and purified as described in Section IV.B.5. and in Section IV.B.6., respectively. The fusion polypeptide SOD-NANB$_{5-1-1}$ was synthesized and purified as described in Section IV.B.1. and in Section IV.D.1., respectively. Purified anti-NANB$_{5-1-1}$ antibodies were obtained as described in Section IV.E.

One hundred microliter aliquots containing varying amounts of purified C100-3 antigen in 0.125M Na borate buffer, pH 8.3, 0.075M NaCl (BBS) was added to each well of a microtiter plate (Dynatech Immulon 2 Removawell Strips). The plate was incubated at 4°C overnight in a humid chamber, after which, the protein solution was removed and the wells washed 3 times with BBS containing 0.02% Triton X-100 (BBST). To prevent non-specific binding, the wells were coated with BSA by addition of 100 microliters of a 5 mg/ml solution of BSA in BBS followed by incubation at room temperature for 1 hour, after which the excess BSA solution was removed. The polypeptides in the coated wells were reacted with purified anti-NANB$_{5-1-1}$ antibodies by adding 1 microgram antibody/well, and incubating the samples for 1 hr at 37°C. After incubation, the excess solution was removed by aspiration, and the wells were washed 5 times with BBST. Anti-NANB$_{5-1-1}$ bound to the fusion polypeptides was determined by the binding of ^{125}I-labeled F'(ab)$_2$ sheep anti-human IgG to the coated wells. Aliquots of 100 microliters of the labeled probe (specific activity 5-20 microcuries/microgram) were added to each well, and the plates were incubated at 37°C for 1 hour, followed by removal of excess probe by aspiration, and 5 washes with BBST. The amount of radioactivity bound in each well was determined by counting in a counter which detects gamma radiation.

The results of the immunological reactivity of C100 with purified anti-$NANB_{5-1-1}$ as compared to that of $NANB_{5-1-1}$ with the purified antibodies are shown in Table 3.

Table 3
Immunological Reactivity of C100-3 compared to $NANB_{5-1-1}$ by Radioimmunoassay

	RIA (cpm/assay)					
AG(ng)	400	320	240	160	60	0
$NANB_{5-1-1}$	7332	6732	4954	4050	3051	57
C100-3	7450	6985	5920	5593	4096	67

The results in Table 3 show that anti-$NANB_{5-1-1}$ recognizes an epitope(s) in the C100 moiety of the C100-3 polypeptide. Thus $NANB_{5-1-1}$ and C100 share a common epitope(s). The results suggest that the cDNA sequence encoding this NANBV epitope(s) is one which is present in both clone 5-1-1 and in clone 81.

IV.H. <u>Characterization of HCV</u>

IV.H.1. <u>Characterization of the Strandedness of the HCV Genome.</u>

The HCV genome was characterized with respect to its strandedness by isolating the nucleic acid fraction from particles captured on anti-$NANB_{5-1-1}$ antibody coated polystyrene beads, and determining whether the isolated nucleic acid hybridized with plus and/or minus strands of HCV cDNA.

Particles were captured from HCV infected chimpanzee plasma using polystyrene beads coated with

immunopurified anti-$NANB_{5-1-1}$ antibody as described in Section IV.F.1. The nucleic acid component of the particles was released using the method described in Section IV.F.2. Aliquots of the isolated genomic nucleic acid equivalent to 3 mls of high titer plasma were blotted onto nitrocellulose filters. As controls, aliquots of denatured HCV cDNA from clone 81 (2 picograms) was also blotted onto the same filters. The filters were probed with ^{32}P-labeled mixture of plus or mixture of minus strands of single stranded DNA cloned from HCV cDNAs; the cDNAs were excised from clones 40b, 81, and 25c.

The single stranded probes were obtained by excising the HCV cDNAs from clones 81, 40b, and 25c with EcoRI, and cloning the cDNA fragments in M13 vectors, mp18 and mp19 [Messing (1983)]. The M13 clones were sequenced to determine whether they contained the plus or minus strands of DNA derived from the HCV cDNAs. Sequencing was by the dideoxychain termination method of Sanger et al. (1977).

Each of a set of duplicate filters containing aliquots of the HCV genome isolated from the captured particles was hybridized with either plus or minus strand probes derived from the HCV cDNAs. Fig. 41 shows the autoradiographs obtained from probing the NANBV genome with the mixture of probes derived from clones 81, 40b, and 25c. This mixture was used to increase the sensitivity of the hybridization assay. The samples in panel I were hybridized with the plus strand probe mixture. The samples in panel II were probed by hybridization with the minus strand probe mixture. The composition of the samples in the panels of the immunoblot are presented in table 4.

Table 4

lane	A	B
1	HCV genome	*
2	----	*
3	*	cDNA 81
4	----	cDNA 81

* is an undescribed sample.

As seen from the results in Fig. 41, only the minus strand DNA probe hybridizes with the isolated HCV genome. This result, in combination with the result showing that the genome is sensitive to RNase and not DNase (See Section IV.C.2.), suggests that the genome of NANBV is positive stranded RNA.

These data, and data from other laboratories concerning the physicochemical properties of a putative NANBV(s), are consistent with the possibility that HCV is a member of the Flaviviridae. However, the possibility that HCV represents a new class of viral agent has not been eliminated.

IV.H.2. Detection of Sequences in Captured Particles Which When Amplified by PCR Hybridize to HCV cDNA Derived from Clone 81

The RNA in captured particles was obtained as described in Section IV.H.1. The analysis for sequences which hybridize to the HCV cDNA derived from clone 81 was

carried out utilizing the PCR amplification procedure, as described in Section IV.C.3, except that the hybridization probe was a kinased oligonucleotide derived from the clone 81 cDNA sequence. The results showed that the amplified sequences hybridized with the clone 81 derived HCV cDNA probe.

IV.H.3. Homology Between the Non-Structural Protein of Dengue Flavivirus (MNWWVD1) and the HCV Polypeptides Encoded by the Combined ORF of Clones 14i Through 39c

The combined HCV cDNAs of clones 14i through 39c contain one continuous ORF, as shown in Fig. 26. The polypeptide encoded therein was analyzed for sequence homology with the region of the non-structural polypeptide(s) in Dengue flavivirus (MNWVD1). The analysis used the Dayhoff protein data base, and was performed on a computer. The results are shown in Fig. 42, where the symbol (:) indicates an exact homology, and the symbol (.) indicates a conservative replacement in the sequence; the dashes indicate spaces inserted into the sequence to achieve the greatest homologies. As seen from the figure, there is significant homology between the sequence encoded in the HCV cDNA, and the non-structural polypeptide(s) of Dengue virus. In addition to the homology shown in Fig. 42, analysis of the polypeptide segment encoded in a region towards the 3'-end of the cDNA also contained sequences which are homologous to sequences in the Dengue polymerase. Of consequence is the finding that the canonical Gly-Asp-Asp (GDD) sequence thought to be essential for RNA-dependent RNA polymerases is contained in the polypeptide encoded in HCV cDNA, in a location which is consistent with that in Dengue 2 virus. (Data not shown.)

IV.H.4. <u>HCV-DNA is Not Detectable in NANBH Infected Tissue</u>

Two types of studies provide results suggesting that HCV-DNA is not detectable in tissue from an individual with NANBH. These results, in conjunction with those described in IV.C. and IV.H.1. and IV.H.2. provide evidence that HCV is not a DNA containing virus, and that its replication does not involve cDNA.

IV.H.4.a. <u>Southern Blotting Procedure</u>

In order to determine whether NANBH infected chimpanzee liver contains detectable HCV-DNA (or HCV-cDNA), restriction enzyme fragments of DNA isolated from this source was Southern blotted, and the blots probed with ^{32}P-labeled HCV cDNA. The results showed that the labeled HCV cDNA did not hybridize to the blotted DNA from the infected chimpanzee liver. It also did not hybridize to control blotted DNA from normal chimpanzee liver. In contrast, in a positive control, a labeled probe of the beta-interferon gene hybridized strongly to Southern blots of restriction enzyme digested human placental DNA. These systems were designed to detect a single copy of the gene which was to be detected with the labeled probe.

DNAs were isolated from the livers of two chimpanzees with NANBH. Control DNAs were isolated from uninfected chimpanzee liver, and from human placentas. The procedure for extracting DNA was essentially according to Maniatis et al. (1982), and the DNA samples were treated with RNAse during the isolation procedure.

Each DNA sample was treated with either EcoRI, MboI, or HincII (12 micrograms), according to the manufacturer's directions. The digested DNAs were electrophoresed on 1% neutral agarose gels, Southern blotted onto nitrocellulose, and the blotted material hybridized with the appropriate nick-translated probe cDNA (3×10^6 cpm/ml of hybridization mix). The DNA from

infected chimpanzee liver and normal liver were hybridized with ^{32}P-labeled HCV cDNA from clones 36 plus 81; the DNA from human placenta was hybridized with ^{32}P-labeled DNA from the beta-interferon gene. After hybridization, the blots were washed under stringent conditions, i.e., with a solution containing 0.1 x SSC, 0.1% SDS, at 65°C.

The beta-interferon gene DNA was prepared as described by Houghton et al (1981).

IV.H.4.b. Amplification by the PCR Technique

In order to determine whether HCV-DNA could be detected in liver from chimpanzees with NANBH, DNA was isolated from the tissue, and subjected to the PCR amplification-detection technique using primers and probe polynucleotides derived from HCV cDNA from clone 81. Negative controls were DNA samples isolated from uninfected HepG2 tissue culture cells, and from presumably uninfected human placenta. Positive controls were samples of the negative control DNAs to which a known relatively small amount (250 molecules) of the HCV cDNA insert from clone 81 was added.

In addition, to confirm that RNA fractions isolated from the same livers of chimpanzees with NANBH contained sequences complementary to the HCV-cDNA probe, the PCR amplification-detection system was also used on the isolated RNA samples.

In the studies, the DNAs were isolated by the procedure described in Section IV.H.4.a, and RNAs were extracted essentially as described by Chirgwin et al. (1981).

Samples of DNA were isolated from 2 infected chimpanzee livers, from uninfected HepG2 cells, and from human placenta. One microgram of each DNA was digested with HindIII according to the manufacturer's directions. The digested samples were subjected to PCR amplification

and detection for amplified HCV cDNA essentially as described in Section IV.C.3., except that the reverse transcriptase step was omitted. The PCR primers and probe were from HCV cDNA clone 81, and are described in Section IV.C.3.. Prior to the amplification, for positive controls, a one microgram sample of each DNA was "spiked" by the addition of 250 molecules of HCV cDNA insert isolated from clone 81.

In order to determine whether HCV sequences were present in RNA isolated from the livers of chimpanzees with NANBH, samples containing 0.4 micrograms of total RNA were subjected to the amplification procedure essentially as described in Section IV.C.3., except that the reverse transcriptase was omitted from some of the samples as a negative control. The PCR primers and probe were from HCV cDNA clone 81, as described supra.

The results showed that amplified sequences complementary to the HCV cDNA probe were not detectable in the DNAs from infected chimpanzee liver, nor were they detectable in the negative controls. In contrast, when the samples, including the DNA from infected chimpanzee liver, was spiked with the HCV cDNA prior to amplification, the clone 81 sequences were detected in all positive control samples. In addition, in the RNA studies, amplified HCV cDNA clone 81 sequences were detected only when reverse transcriptase was used, suggesting strongly that the results were not due to a DNA contamination.

These results show that hepatocytes from chimpanzees with NANBH contain no, or undetectable levels, of HCV DNA. Based upon the spiking study, if HCV DNA is present, it is at a level far below .06 copies per hepatocyte. In contrast, the HCV sequences in total RNA from the same liver samples was readily detected with the PCR technique.

IV.I. <u>ELISA Determinations for HCV Infection Using HCV c100-3 As Test Antigen</u>

All samples were assayed using the HCV c100-3 ELISA. This assay utilizes the HCV c100-3 antigen (which was synthesized and purified as described in Section IV.B.5), and a horseradish peroxidase (HRP) conjugate of mouse monoclonal anti-human IgG.

Plates coated with the HCV c100-3 antigen were prepared as follows. A solution containing Coating buffer (50mM Na Borate, pH 9.0), 21 ml/plate, BSA (25 micrograms/ml), c100-3 (2.50 micrograms/ml) was prepared just prior to addition to the Removeawell Immulon I plates (Dynatech Corp.). After mixing for 5 minutes, 0.2ml/well of the solution was added to the plates, they were covered and incubated for 2 hours at 37°C, after which the solution was removed by aspiration. The wells were washed once with 400 microliters Wash Buffer (100 mM sodium phosphate, pH 7.4, 140 mM sodium chloride, 0.1% (W/V) casein, 1% (W/V) Triton x-100, 0.01% (W/V) Thimerosal). After removal of the wash solution, 200 microliters/well of Postcoat solution (10 mM sodium phosphate, pH 7.2, 150 mM sodium chloride, 0.1% (w/v) casein and 2 mM phenylmethylsulfonylfluoride (PMSF)) was added, the plates were loosely covered to prevent evaporation, and were allowed to stand at room temperature for 30 minutes. The wells were then aspirated to remove the solution, and lyophilized dry overnight, without shelf heating. The prepared plates may be stored at 2-8°C in sealed aluminum pouches.

In order to perform the ELISA determination, 20 microliters of serum sample or control sample was added to a well containing 200 microliters of sample diluent (100 mM sodium phosphate, pH 7.4, 500 mM sodium chloride, 1 mM EDTA, 0.1% (W/V) Casein, 0.015 (W/V) Therosal, 1% (W/V) Triton X-100, 100 micrograms/ml yeast extract). The

plates were sealed, and incubated at 37°C for two hours, after which the solution was removed by aspiration, and the wells were washed with 400 microliters of wash buffer (phosphate buffered saline (PBS) containing 0.05% Tween 20). The washed wells were treated with 200 microliters of mouse anti-human IgG-HRP conjugate contained in a solution of Ortho conjugate diluent (10 mM sodium phosphate, pH 7.2, 150 mM sodium chloride, 50% (V/V) fetal bovine serum, 1% (V/V) heat treated horse serum, 1 mM $K_3Fe(CN)_6$, 0.05% (W/V) Tween 20, 0.02% (W/V) Thimerosal). Treatment was for 1 hour at 37°C, the solution was removed by aspiration, and the wells were washed with wash buffer, which was also removed by aspiration. To determine the amount of bound enzyme conjugate, 200 microliters of substrate solution (10 mg O-phenylenediamine dihydrochloride per 5 ml of Developer solution) was added. Developer solution contains 50 mM sodium citrate adjusted to pH 5.1 with phosphoric acid, and 0.6 microliters/ml of 30% H_2O_2. The plates containing the substrate solution were incubated in the dark for 30 minutes at room temperature, the reactions were stopped by the addition of 50 microliters/ml 4N sulfuric acid, and the ODs determined.

The examples provided below show that the microtiter plate screening ELISA which utilizes HCV c100-3 antigen has a high degree of specificity, as evidenced by an initial rate of reactivity of about 1%, with a repeat reactive rate of about 0.5% on random donors. The assay is capable of detecting an immunoresponse in both the post acute phase of the infection, and during the chronic phase of the disease. In addition, the assay is capable of detecting some samples which score negative in the surrogate tests for NANBH; these samples come from individuals with a history of NANBH, or from donors implicated in NANBH transmission.

In the examples described below, the following abbreviations are used:

ALT	Alanine amino transferase
Anti-HBc	Antibody against HBc
Anti-HBsAg	Antibody against HBsAg
HBc	Hepatitis B core antigen
ABsAg	Hepatitis B surface antigen
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IU/L	International units/Liter
NA	Not available
NT	Not tested
N	Sample size
Neg	Negative
OD	Optical density
Pos	Positive
S/CO	Signal/cutoff
SD	Standard deviation
x	Average or mean
WNL	Within normal limits

IV.I.1. <u>HCV Infection in a Population of Random Blood Donors</u>

A group of 1,056 samples (fresh sera) from random blood donors were obtained from Irwin Memorial Blood Bank, San Francisco, California. The test results obtained with these samples are summarized in a histogram showing the distribution of the OD values (Fig. 43). As seen in Fig. 43, 4 samples read >3, 1 sample reads between 1 and 3, 5 samples read between 0.4 and 1, and the remaining 1,046 samples read <0.4, with over 90% of these samples reading <0.1.

The results on the reactive random samples are presented in Table 5. Using a cut-off value equal to the

mean plus 5 standard deviations, ten samples out of the 1,056 (0.95%) were initially reactive. Of these, five samples (0.47%) repeated as reactive when they were assayed a second time using the ELISA. Table 5 also shows the ALT and Anti-HBd status for each of the repeatedly reactive samples. Of particular interest is the fact that all five repeat reactive samples were negative in both surrogate tests for NANBH, while scoring positive in the HCV ELISA.

TABLE 5

RESULTS ON REACTIVE RANDOM SAMPLES

N = 1051
$\bar{x}$ = 0.049*
SD = ± 0.074
Cut-off: $\bar{x}$ + 5SD = 0.419 (0.400 + Negative Control)

Samples	Initial Reactives OD	Repeat Reactives OD	ALT** (IU/L)	Anti HBc*** (OD)
4227	0.462	0.084	NA	NA
6292	0.569	0.294	NA	NA
6188	0.699	0.326	NA	NA
6157	0.735	0.187	NA	NA
6277	0.883	0.152	NA	NA
6397	1.567	1.392	30.14	1.433
6019	>3.000	>3.000	46.48	1.057
6651	>3.000	>3.000	48.53	1.343
6669	>3.000	>3.000	60.53	1.165
4003	>3.000	3.000	WNL****	Negative
	10/1056 = 0.95%	5/1056 = 0.47%		

* Samples reading >1.5 were not included in calculating the Mean and SD

** ALT ≥ 68 IU/L is above normal limits.

*** Anti-HBc ≤ 0.535 (competition assay) is considered positive.

**** WNL: Within normal limits.

IV.I.2. Chimpanzee Serum Samples

Serum samples from eleven chimpanzees were tested with the HCV c100-3 ELISA. Four of these chimpanzees were infected with NANBH from a contaminated batch of Factor VIII (presumably Hutchinson strain), following an established procedure in a collaboration with Dr. Daniel Bradley at the Centers for Disease Control. As controls, four other chimpanzees were infected with HAV and three with HBV. Serum samples were obtained at different times after infection.

The results, which are summarized in Table 6, show documented antibody seroconversion in all chimpanzees infected with the Hutchinson strain of NANBH. Following the acute phase of infection (as evidenced by the significant rise and subsequent return to normal of ALT levels), antibodies to HCV c100-3 became detectable in the sera of the 4/4 NANBH infected chimpanzees. These samples had previously been shown, as discussed in Section IV.B.3., to be positive by a Western analysis, and an RIA. In contrast, none of the control chimpanzees which had been infected with HAV or HBV showed evidence of reactivity in the ELISA.

TABLE 6
CHIMPANZEE SERUM SAMPLES

	OD	S/CO	Inoculation Date	Bleed Date	ALT (IU/L)	Transfused
Negative Control	0.001					
Positive Control	1.504					
Cutoff	0.401					
Chimp 1	-0.007	0.00	05/24/84	05/24/84	9	NANB
	0.003	0.01		08/07/84	71	
	>3.000	>7.48		09/18/84	19	
	>3.000	>7.48		10/24/84	---	
Chimp 2	---	---	06/07/84	---	---	NANB
	-0.003	0.00		05/31/84	5	
	-0.005	0.00		06/28/84	52	
	0.945	2.36		08/20/84	13	
	>3.000	>7.48		10/24/84	---	
Chimp 3	0.005	0.01	03/14/85	03/14/85	8	NANB
	0.017	0.04		04/26/85	205	
	0.006	0.01		05/06/85	14	
	1.010	2.52		08/20/85	6	
Chimp 4	-0.006	0.00	03/11/85	03/11/85	11	NANB
	0.003	0.01		05/09/85	132	
	0.523	1.31		06/06/85	---	
	1.574	3.93		08/01/85	---	
Chimp 5	-0.006	0.00	11/21/80	11/21/80	4	HAV
	0.001	0.00		12/16/80	147	
	0.003	0.01		12/30/80	18	
	0.006	0.01		07/29 - 08/21/81	5	
Chimp 6	---	---	05/25/82	---	---	HAV
	-0.005	0.00		05/17/82	---	
	0.001	0.00		06/10/82	106	
	-0.004	0.00		07/06/82	10	
	0.290	0.72		10/01/82	---	
Chimp 7	-0.008	0.00	05/25/82	05/25/82	7	HAV
	-0.004	0.00		06/17/82	83	
	-0.006	0.00		09/16/82	5	
	0.005	0.01		10/09/82	---	

TABLE 6

CHIMPANZEE SERUM SAMPLES

(Cont'd)

	OD	S/CO	Inoculation Date	Bleed Date	ALT (IU/L)	Transfused
Chimp 8	-0.007	0.00	11/21/80	11/21/80	15	HAV
	0.000	0.00		12/16/80	130	
	0.004	0.01		02/03/81	8	
	0.000	0.00		06/03 - 06/10/81	4.5	
Chimp 9	---	---	07/24/80	---	---	HBV
	0.019	0.05		08/22 - 10/10/79	---	
	---	---		03/11/81	57	
	0.015	0.04		07/01 - 08/05/81	9	
	0.008	0.02		10/01/81	6	
Chimp 10	---	---	05/12/82	---	---	HBV
	0.011	0.03		04/21 - 05/12/82	9	
	0.015	0.04		09/01 - 09/08/82	126	
	0.008	0.02		12/02/82	9	
	0.010	0.02		01/06/83	13	
Chimp 11	---	---	05/12/82	---	---	HBV
	0.000	0.00		01/06 - 05/12/82	11	
	---	---		06/23/82	100	
	-0.003	0.00		06/09 - 07/07/82	---	
	-0.003	0.00		10/28/82	9	
	-0.003	0.00		12/20/82	10	

IV.I.3. <u>Panel 1: Proven Infectious Sera from Chronic Human NANBH Carriers</u>

A coded panel consisted of 22 unique samples, each one in duplicate, for a total of 44 samples. The samples were from proven infectious sera from chronic NANBH carriers, infectious sera from implicated donors, and infectious sera from acute phase NANBH patients. In addition, the samples were from highly pedigreed negative controls, and other disease controls. This panel was provided by Dr. H. Alter of the Department of Health and Human Services, National Institutes of Health, Bethesda, Maryland. The panel was constructed by Dr. Alter several years ago, and has been used by Dr. Alter as a qualifying panel for putative NANBH assays.

The entire panel was assayed twice with the ELISA assay, and the results were sent to Dr. Alter to be scored. The results of the scoring are shown in Table 7. Although the Table reports the results of only one set of duplicates, the same values were obtained for each of the duplicate samples.

As shown in Table 7, 6 sera which were proven infectious in a chimpanzee model were strongly positive. The seventh infectious serum corresponded to a sample for an acute NANBH case, and was not reactive in this ELISA. A sample from an implicated donor with both normal ALT levels and equivocal results in the chimpanzee studies was non-reactive in the assay. Three other serial samples from one individual with acute NANBH were also non-reactive. All samples coming from the highly pedigreed negative controls, obtained from donors who had at least 10 blood donations without hepatitis implication, were non-reactive in the ELISA. Finally, four of the samples tested had previously scored as positive in putative NANBH assays developed by others, but these assays were not

confirmable. These four samples scored negatively with the HCV ELISA.

TABLE 7

H. ALTER'S PANEL 1:

Panel	1st Result	2nd Result
1) Proven Infectious by Chimpanzee Transmission		
A. Chronic NANB: Post-Tx		
JF	+	+
EB	+	+
PG	+	+
B. Implicated Donors with Elevated ALT		
BC	+	+
JJ	+	+
BB	+	+
C. Acute NANB: Post-Tx		
WH	-	-
2) Equivocally Infectious by Chimpanzee Transmission		
A. Implicated Donor with Normal ALT		
CC	-	-
3) Acute NANB: Post-Tx		
JL Week 1	-	-
JL Week 2	-	-
JL Week 3	-	-
4) Disease Controls		
A. Primary Biliary Cirrhosis		
EK	-	-
B. Alcoholic Hepatitis in Recovery		
HB	-	-
5) Pedigreed Negative Controls		
DH	-	-
DC	-	-
LV	-	-
HL	-	-
AH	-	-
6) Potential NANB "Antigens"		
JS-80-01T-0 (Ishida)	-	-
Asterix (Trepo)	-	-
Zurtz (Arnold)	-	-
Becassdine (Trepo)	-	-

IV.I.4. <u>Panel 2: Donor/Recipient NANBH</u>

The coded panel consisted of 10 unequivocal donor-recipient cases of transfusion associated NANBH, with a total of 188 samples. Each case consisted of samples of some or all the donors to the recipient, and of serial samples (drawn 3, 6, and 12 months after transfusion) from the recipient. Also included was a pre-bleed, drawn from the recipient before transfusion. The coded panel was provided by Dr. H. Alter, from the NIH, and the results were sent to him for scoring.

The results, which are summarized in Table 8, show that the ELISA detected antibody seroconversion in 9 of 10 cases of transfusion associated NANBH. Samples from case 4 (where no seroconversion was detected), consistently reacted poorly in the ELISA. Two of the 10 recipient samples were reactive at 3 months post transfusion. At six months, 8 recipient samples were reactive; and at twelve months, with the exception of case 4, all samples were reactive. In addition, at least one antibody positive donor was found in 7 out of the 10 cases, with case 10 having two positive donors. Also, in case 10, the recipient's pre-bleed was positive for HCV antibodies. The one month bleed from this recipient dropped to borderline reactive levels, while it was elevated to positive at 4 and 10 month bleeds. Generally, a S/CO of 0.4 is considered positive. Thus, this case may represent a prior infection of the individual with HCV.

The ALT and HBc status for all the reactive, i.e., positive, samples are summarized in Table 9. As seen in the table, 1/8 donor samples was negative for the surrogate markers and reactive in the HCV antibody ELISA. On the other hand, the recipient samples (followed up to 12 months after transfusion) had either elevated ALT, positive Anti-HBc, or both.

TABLE 8

DONOR/RECIPIENT NANB PANEL

H. ALTER DONOR/RECIPIENT NANB PANEL

Case	Donor		Recipient Prebleed		3 Months		Post-TX 6 Months		12 Months	
	OD	S/CO	OD	S/CO	OD	S/CO	OD	S/CO	OD	S/CO
1.	---	---	.032	0.07	.112	0.26	>3.000	>6.96	>3.000	>6.96
2.	---	---	.059	0.14	.050	0.12	1.681	3.90	>3.000	>6.96
3.	.403	0.94	.049	0.11	.057	0.13	>3.000	>6.96	>3.000	>6.96
4.	---	---	.065	0.15	.073	0.17	.067	0.16	.217	0.50
5.	>3.000	>6.96	.034	0.08	.096	0.22	>3.000	>6.96	>3.000	>6.96
6.	>3.000	>6.96	.056	0.13	1.475	3.44	>3.000	>6.96	>3.000	>6.96
7.	>3.000	>6.96	.034	0.08	.056	0.13	>3.000	>6.96	>3.000	>6.96
8.	>3.000	>6.96	.061	0.14	.078	0.18	2.262	5.28	>3.000	>6.96
9.	>3.000	>6.96	.080	0.19	.127	0.30	.055	0.13	>3.000	>6.96
10.	>3.000	>6.96	>3.000	>6.96	.317*	0.74	>3.000**	>6.96	>3.000***	>6.96
	>3.000	>6.96								

* 1 MONTH, ** 4 MONTHS, *** 10 MONTHS

TABLE 9

ALT AND HBc STATUS FOR REACTIVE SAMPLES IN H. ALTER PANEL 1

Samples		Anti- ALT*	HBc**
Donors			
Case 3		Normal	Negative
Case 5		Elevated	Positive
Case 6		Elevated	Positive
Case 7		Not available	Negative
Case 8		Normal	Positive
Case 9		Elevated	Not available
Case 10		Normal	Positive
Case 10		Normal	Positive
Recipients			
Case 1	6 mo	Elevated	Positive
12 mo	Elevated	Not tested	
Case 2	6 mo	Elevated	Negative
12 mo	Elevated	Not tested	
Case 3	6 mo	Normal	Not tested***
12 mo	Elevated	Not tested***	
Case 5	6 mo	Elevated	Not tested
12 mo	Elevated	Not tested	
Case 6	3 mo	Elevated	Negative
6 mo	Elevated	Negative	
12 mo	Elevated	Not tested	
Case 7	6 mo	Elevated	Negative
12 mo	Elevated	Negative	
Case 8	6 mo	Normal	Positive
12 mo	Elevated	Not tested	
Case 9	12 mo	Elevated	Not tested
Case 10	4 mo	Elevated	Not tested
10 mo	Elevated	Not tested	

* ALT ≥45 IU/L is above normal limits.
** Anti-HBc ≤50% (competition assay) is considered positive.
*** Prebleed and 3 mo samples were negative for HBc.

IV.I.5. Determination of HCV Infection in High Risk Group Samples

Samples from high risk groups were monitored using the ELISA to determine reactivity to HCV c100-3 antigen. These samples were obtained from Dr. Gary Tegtmeier, Community Blood Bank, Kansas City. The results are summarized in Table 10.

As shown in the table, the samples with the highest reactivity are obtained from hemophiliacs (76%). In addition, samples from individuals with elevated ALT and positive for Anti-HBc, scored 51% reactive, a value which is consistent with the value expected from clinical data and NANBH prevalence in this group. The incidence of antibody to HCV was also higher in blood donors with elevated ALT alone, blood donors positive for antibodies to Hepatitis B core alone, and in blood donors rejected for reasons other than high ALT or anti-core antibody when compared to random volunteer donors.

TABLE 10

NANBH HIGH RISK GROUP SAMPLES

Group		N	Distribution N	Distribution OD	% Reactive
Elevated ALT		35	3	>3.000	11.4%
	1	0.728			
Anti-HBc		24	5	>3.000	20.8%
Elevated ALT, Anti-HBc		33	12	>3.000	51.5%
	1	2.768			
	1	2.324			
	1	0.939			
	1	0.951			
	1	0.906			
Rejected Donors		25	5	>3.000	20.0%
Donors with History of Hepatitis		150	19	>3.000	14.7%
	1	0.837			
	1	0.714			
	1	0.469			
Haemophiliacs		50	31	>3.000	76.0%
	1	2.568			
	1	2.483			
	1	2.000			
	1	1.979			
	1	1.495			
	1	1.209			
	1	0.819			

IV.I.6 Comparative Studies Using Anti-IgG or Anti-IgM Monoclonal Antibodies, or Polyclonal Antibodies as a Second Antibody in the HCV c100-3 ELISA

The sensitivity of the ELISA determination which uses the anti-IgG monoclonal conjugate was compared to that obtained by using either an anti-IgM monoclonal conjugate, or by replacing both with a polyclonal antiserum reported to be both heavy and light chain specific. The following studies were performed.

IV.I.6.a. Serial Samples from Seroconverters

Serial samples from three cases of NANB seroconverters were studied in the HCV c100-3 ELISA assay using in the enzyme conjugate either the anti-IgG monoclonal alone, or in combination with an anti-IgM monoclonal, or using a polyclonal antiserum. The samples were provided by Dr. Cladd Stevens, N.Y. Blood Center, N.Y.C., N.Y.. The sample histories are shown in Table 11.

The results obtained using an anti-IgG monoclonal antibody-enzyme conjugate are shown in Table 12. The data shows that strong reactivity is initially detected in samples 1-4, 2-8, and 3-5, of cases 1, 2, and 3, respectively.

The results obtained using a combination of an anti-IgG monoclonal conjugate and an anti-IgM conjugate are shown in Table 13. Three different ratios of anti-IgG to anti-IgM were tested; the 1:10,000 dilution of anti-IgG was constant throughout. Dilutions tested for the anti-IgM monoclonal conjugate were 1:30,000, 1:60,000, and 1:120,000. The data shows that, in agreement with the studies with anti-IgG alone, initial strong reactivity is detected in samples 1-4, 2-8, and 3-5.

The results obtained with the ELISA using anti-IgG monoclonal conjugate (1:10,000 dilution), or Tago

polyclonal conjugate (1:80,000 dilution), or Jackson polyclonal conjugate (1:80,000 dilution) are shown in Table 14. The data indicates that initial strong reactivity is detected in samples 1-4, 2-8, and 3-5 using all three configurations; the Tago polyclonal antibodies yielded the lowest signals.

The results presented above show that all three configurations detect reactive samples at the same time after the acute phase of the disease (as evidenced by the ALT elevation). Moreover, the results indicate that the sensitivity of the HCV c100-3 ELISA using anti-IgG monoclonal-enzyme conjugate is equal to or better than that obtained using the other tested configurations for the enzyme conjugate.

TABLE 11

DESCRIPTION OF SAMPLES FROM CLADD STEVENS PANEL

	Date	HBsAg	Anti-HBs	Anti-HBc	ALT	Bilirubin
Case 1						
1-1	8/5/81	1.0	91.7	12.9	40.0	-1.0
1-2	9/2/81	1.0	121.0	15.1	274.0	1.4
1-3	10/7/81	1.0	64.0	23.8	261.0	0.9
1-4	11/19/81	1.0	67.3	33.8	75.0	0.9
1-5	12/15/81	1.0	50.5	27.6	71.0	1.0
Case 2						
2-1	10/19/81	1.0	1.0	116.2	17.0	-1.0
2-2	11/17/81	1.0	0.8	89.5	46.0	1.1
2-3	12/02/81	1.0	1.2	78.3	63.0	1.4
2-4	12/14/81	1.0	0.9	90.6	152.0	1.4
2-5	12/23/81	1.0	0.8	93.6	624.0	1.7
2-6	1/20/82	1.0	0.8	92.9	66.0	1.5
2-7	2/15/82	1.0	0.8	86.7	70.0	1.3
2-8	3/17/82	1.0	0.9	69.8	24.0	-1.0
2-9	4/21/82	1.0	0.9	67.1	53.0	1.5
2-10	5/19/82	1.0	0.5	74.8	95.0	1.6
2-11	6/14/82	1.0	0.8	82.9	37.0	-1.0
Case 3						
3-1	4/7/81	1.0	1.2	88.4	13.0	-1.0
3-2	5/12/81	1.0	1.1	126.2	236.0	0.4
3-3	5/30/81	1.0	0.7	99.9	471.0	0.2
3-4	6/9/81	1.0	1.2	110.8	315.0	0.4
3-5	7/6/81	1.0	1.1	89.9	273.0	0.4
3-6	8/10/81	1.0	1.0	118.2	158.0	0.4
3-7	9/8/81	1.0	1.0	112.3	84.0	0.3
3-8	10/14/81	1.0	0.9	102.5	180.0	0.5
3-9	11/11/81	1.0	1.0	84.6	154.0	0.3

TABLE 12

ELISA RESULTS OBTAINED USING AN ANTI-IgG MONOCLONAL CONJUGATE

SAMPLE	DATE	ALT	OD	S/CO
NEG CONTROL			.076	
CUTOFF			.476	
PC (1:128)			1.390	
CASE #1				
1-1	08/05/81	40.0	.178	.37
1-2	09/02/81	274.0	.154	.32
1-3	10/07/81	261.0	.129	.27
1-4	11/19/81	75.0	.937	1.97
1-5	12/15/81	71.0	>3.000	>6.30
CASE #2				
2-1	10/19/81	17.0	.058	0.12
2-2	11/17/81	46.0	.050	0.11
2-3	12/02/81	63.0	.047	0.10
2-4	12/14/81	152.0	.059	0.12
2-5	12/23/81	624.0	.070	0.15
2-6	01/20/82	66.0	.051	0.11
2-7	02/15/82	70.0	.139	0.29
2-8	03/17/82	24.0	1.867	3.92
2-9	04/21/82	53.0	>3.000	>6.30
2-10	05/19/82	95.0	>3.000	>6.30
2-11	06/14/82	37.0	>3.000	>6.30
CASE #3				
3-1	04/07/81	13.0	.090	.19
3-2	05/12/81	236.0	.064	.13
3-3	05/30/81	471.0	.079	.17
3-4	06/09/81	315.0	.211	.44
3-5	07/06/81	273.0	1.707	3.59
3-6	08/10/81	158.0	>3.000	>6.30
3-7	09/08/81	84.0	>3.000	>6.30
3-8	10/14/81	180.0	>3.000	>6.30
3-9	11/11/81	154.0	>3.000	>6.30

TABLE 13

ELISA RESULTS OBTAINED USING ANTI-IgG and ANTI-IgM MONOCLONAL CONJUGATE

			NANB ELISAs					
			Monoclonals IgG 1:10K IgM 1:30K		Monoclonals IgG 1:10K IgM 1:60K		Monoclonals IgG 1:10K IgM 1:120K	
SAMPLE	DATE	ALT	OD	S/CO	OD	S/CO	OD	S/CO
Neg Control			.100		.080		.079	
Cutoff								
PC (1:128)			1.083		1.328		1.197	
Case #1								
1-1	08/05/81	40	.173		.162		.070	
1-2	09/02/81	274	.194		.141		.079	
1-3	10/07/81	261	.162		.129		.063	
1-4	11/19/81	75	.812		.85		.709	
1-5	12/15/81	71	>3.00		>3.00		>3.00	
Case #2								
2-1	10/19/81	17	.442		.045		.085	
2-2	11/17/81	46	.102		.029		.030	
2-3	12/02/81	63	.059		.036		.027	
2-4	12/14/81	152	.065		.041		.025	
2-5	12/23/81	624	.082		.033		.032	
2-6	01/20/82	66	.102		.042		.027	
2-7	02/15/82	70	.188		.068		.096	
2-8	03/17/82	24	1.728		1.668		1.541	
2-9	04/21/82	53	>3.00		2.443		>3.00	
2-10	05/19/82	95	>3.00		>3.00		>3.00	
2-11	06/14/82	37	>3.00		>3.00		>3.00	
Case #3								
3-1	04/07/81	13	.193		.076		.049	
3-2	05/12/81	236	.201		.051		.038	
3-3	05/30/81	471	.132		.067		.052	
3-4	06/09/81	315	.175		.155		.140	
3-5	07/06/81	273	1.335		1.238		1.260	
3-6	08/10/81	158	>3.00		>3.00		>3.00	
3-7	09/08/81	84	>3.00		>3.00		>3.00	
3-8	10/14/81	180	>3.00		>3.00		>3.00	
3-9	11/11/81	154	>3.00		>3.00		>3.00	

TABLE 14

ELISA RESULTS OBTAINED USING POLYCLONAL CONJUGATES

			NANB ELISAs					
			MONOCLONAL 1:10K		TAGO 1:80K		JACKSON 1:80K	
SAMPLE	DATE	ALT	OD	S/CO	OD	S/CO	OD	S/CO
NEG CONTROL			.076		.045		.154	
CUTOFF			.476		.545		.654	
PC (1:128)			1.390		.727		2.154	
CASE #1								
1-1	08/05/81	40	.178	.37	.067	.12	.153	.23
1-2	09/02/81	274	.154	.32	.097	.18	.225	.34
1-3	10/07/81	261	.129	.27	.026	.05	.167	.26
1-4	11/19/81	75	.937	1.97	.324	.60	.793	1.21
1-5	12/15/81	71	>3.00	>6.30	1.778	3.27	>3.00	>4.59
CASE #2								
2-1	10/19/81	17	.058	.12	.023	.04	.052	.08
2-2	11/17/81	46	.050	.11	.018	.03	.058	.09
2-3	12/02/81	63	.047	.10	.020	.04	.060	.09
2-4	12/14/81	152	.059	.12	.025	.05	.054	.08
2-5	12/23/81	624	.070	.15	.026	.05	.074	.11
2-6	01/20/82	66	.051	.11	.018	.03	.058	.09
2-7	02/15/82	70	.139	.29	.037	.07	.146	.22
2-8	03/17/82	24	1.867	3.92	.355	.65	1.429	2.19
2-9	04/21/82	53	>3.00	>6.30	.748	1.37	>3.00	>4.59
2-10	05/19/82	95	>3.00	>6.30	1.025	1.88	>3.00	>4.59
2-11	06/14/82	37	>3.00	>6.30	.917	1.68	>3.00	>4.59
CASE #3								
3-1	04/07/81	13	.090	.19	.049	.09	.138	.21
3-2	05/12/81	236	.064	.13	.040	.07	.094	.14
3-3	05/30/81	471	.079	.17	.045	.08	.144	.22
3-4	06/09/81	315	.211	.44	.085	.16	.275	.42
3-5	07/06/81	273	1.707	3.59	.272	.50	1.773	2.71
3-6	08/10/81	158	>3.00	>6.30	1.347	2.47	>3.00	>4.59
3-7	09/08/81	84	>3.00	>6.30	2.294	4.21	>3.00	>4.59
3-8	10/14/81	180	>3.00	>6.30	>3.00	>5.50	>3.00	>4.59
3-9	11/11/81	154	>3.00	>6.30	>3.00	>5.50	>3.00	>4.59

IV.I.6.b. Samples from Random Blood Donors

Samples from random blood donors (See Section IV.I.1.) were screened for HCV infection using the HCV c100-3 ELISA, in which the antibody-enzyme conjugate was either an anti-IgG monoclonal conjugate, or a polyclonal conjugate. The total number of samples screened were 1077 and 1056, for the polyclonal conjugate and the monoclonal conjugate, respectively. A summary of the results of the screening is shown in Table 15, and the sample distributions are shown in the histogram in Fig. 44.

The calculation of the average and standard deviation was performed excluding samples that gave a signal over 1.5, i.e., 1073 OD values were used for the calculations utilizing the polyclonal conjugate, and 1051 for the anti-IgG monoclonal conjugate. As seen in Table 15, when the polyclonal conjugate was used, the average was shifted from 0.0493 to 0.0931, and the standard deviation was increased from 0.074 to 0.0933. Moreover, the results also show that if the criteria of x +5SD is employed to define the assay cutoff, the polyclonal-enzyme conjugate configuration in the ELISA requires a higher cutoff value. This indicates a reduced assay specificity as compared to the monoclonal system. In addition, as depicted in the histogram in Fig. 44, a greater separation of results between negative and positive distributions occurs when random blood donors are screened in an ELISA using the anti-IgG monoclonal conjugate as compared to the assay using a commercial polyclonal label.

TABLE 15

COMPARISON OF TWO ELISA CONFIGURATIONS IN TESTING SAMPLES FROM RANDOM BLOOD DONORS

CONJUGATE	POLYCLONAL (Jackson)	ANTI-IgG MONOCLONAL
Number of samples	1073	1051
Average (x)	0.0931	0.04926
Standard deviation (SD)	0.0933	0.07427
5 SD	0.4666	0.3714
CUT-OFF (5 SD + x)	0.5596	0.4206

IV.J. Detection of HCV Seroconversion in NANBH Patients from a Variety of Geographical Locations

Sera from patients who were suspected to have NANBH based upon elevated ALT levels, and who were negative in HAV and HBV tests were screened using the RIA essentially as described in Section IV.D., except that the HCV C100-3 antigen was used as the screening antigen in the microtiter plates. As seen from the results presented in Table 16, the RIA detected positive samples in a high percentage of the cases.

Table 16

Seroconversion Frequencies for Anti-c100-3 Among NANBH Patients in Different Countries

Country	The Netherlands	Italy	Japan
No. Examined	5	36	26
No. Positive	3	29	19
% Positive	60	80	73

IV.K. Detection of HCV Seroconversion in Patients with "Community Acquired" NANBH

Sera which was obtained from 100 patients with NANBH, for whom there was no obvious transmission route (i.e., no transfusions, i.v. drug use, promiscuity, etc. were identified as risk factors), was provided by Dr. M. Alter of the Center for Disease Control, and Dr. J.

Dienstag of Harvard University. These samples were screened using an RIA essentially as described in Section IV.D., except that the HCV c100-3 antigen was used as the screening antigen attached to the microtiter plates. The results showed that of the 100 serum samples, 55 contained antibodies that reacted immunologically with the HCV c100-3 antigen.

The results described above suggest that "Community Acquired" NANBH is also caused by HCV. Moreover, since it has been demonstrated herein that HCV is related to Flaviviruses, most of which are transmitted by arthropods, it is suggestive that HCV transmission in the "Community Acquired" cases also results from arthropod transmission.

IV.L. <u>Comparison of Incidence of HCV Antibodies and Surrogate Markers in Donors Implicated in NANBH Transmission</u>

A prospective study was carried out to determine whether recipients of blood from suspected NANBH positive donors, who developed NANBH, seroconverted to anti-HCV-antibody positive. The blood donors were tested for the surrogate marker abnormalities which are currently used as markers for NANBH infection, i.e., elevated ALT levels, and the presence of anti-core antibody. In addition, the donors were also tested for the presence of anti-HCV antibodies. The determination of the presence of anti-HCV antibodies was determined using a radioimmunoassay as described in Section IV.K. The results of the study are presented in Table 17, which shows: the patient number (column 1); the presence of anti-HCV antibodies in patient serum (column 2); the number of donations received by the patient, with each donation being from a different donor (column 3); the presence of anti-HCV antibodies in donor serum (column 4); and the surrogate abnormality of the

donor (column 5) (NT or -- means not tested) (ALT is elevated transaminase, and ANTI-HBc is anti-core antibody).

The results in Table 17 demonstrate that the HCV antibody test is more accurate in detecting infected blood donors than are the surrogate marker tests. Nine out of ten patients who developed NANBH symptoms tested positive for anti-HCV antibody seroconversion. Of the 11 suspected donors, (patient 6 received donations from two different individuals suspected of being NANBH carriers), 9 were positive for anti-HCV antibodies, and 1 was borderline positive, and therefore equivocal (donor for patient 1). In contrast, using the elevated ALT test 6 of the ten donors tested negative, and using the anticore-antibody test 5 of the ten donors tested negative. Of greater consequence, though, in three cases (donors to patients 8, 9, and 10) the ALT test and the ANTI-HBc test yielded inconsistent results.

Table 17

DEVELOPMENT OF ANTI-HCV ANTIBODIES IN PATIENTS RECEIVING BLOOD FROM DONORS SUSPECTED OF BEING NANBH CARRIERS

Patient	Anti-HCV Seroconversion in Patient	No. of Donations/Donors	Anti-HCV Positive Donor	Surrogate Abnormality Alt	Surrogate Abnormality Anti-HB
1	yes	18	equiv	no	no
2	yes	18	yes	NT	yes
3	yes	13	yes	no	no
4	no	18	no	--	--
5	yes	16	yes	yes	yes
6	yes	11	yes(2)	no	no
				yes	yes
7	yes	15	yes	NT	no
8	yes	20	yes	no	yes
9	yes	5	yes	yes	no
10	yes	15	yes	no	yes

*Same donor as anti-NANBV Positive.

IV.M. Amplification for Cloning of HCV cDNA Sequences Utilizing the PCR and Primers Derived from Conserved Regions of Flavivirus Genomic Sequences

The results presented supra., which suggest that HCV is a flavivirus or flavi-like virus, allows a strategy for cloning uncharacterized HCV cDNA sequences utilizing the PCR technique, and primers derived from the regions encoding conserved amino acid sequences in flaviviruses. Generally, one of the primers is derived from a defined HCV genomic sequence, and the other primer which flanks a region of unsequenced HCV polynucleotide is derived from a conserved region of the flavivirus genome. The flavivirus genomes are known to contain conserved sequences within the NS1, and E polypeptides, which are encoded in the 5'-region of the flavivirus genome. Corresponding sequences encoding these regions lie upstream of the HCV cDNA sequence shown in Fig. 26. Thus, to isolate cDNA sequences derived from this region of the HCV genome, upstream primers are designed which are derived from the conserved sequences within these flavivirus polypeptides. The downstream primers are derived from an upstream end of the known portion of the HCV cDNA.

Because of the degeneracy of the code, it is probable that there will be mismatches between the flavivirus probes and the corresponding HCV genomic sequence. Therefore a strategy which is similar to the one described by Lee (1988) is used. The Lee procedure utilizes mixed oligonucleotide primers complementary to the reverse translation products of an amino acid sequence; the sequences in the mixed primers takes into account every codon degeneracy for the conserved amino acid sequence.

Three sets of primer mixes are generated, based on the amino acid homologies found in several flaviviruses, including Dengue-2,4 (D-2,4), Japanese

Encephalitis Virus (JEV), Yellow Fever (YF), and West Nile Virus (WN). The primer mixture derived from the most upstream conserved sequence (5'-1), is based upon the amino acid sequence gly-trp-gly, which is part of the conserved sequence asp-arg-gly-trp-gly-aspN found in the E protein of D-2, JEV, YF, and WN. The next primer mixture (5'-2) is based upon a downstream conserved sequence in E protein, phe-asp-gly-asp-ser-tyr-ileu-phe-gly-asp-ser-tyr-ileu, and is derived from phe-gly-asp; the conserved sequence is present in D-2, JEV, YF, and WN. The third primer mixture (5'-3), is based on the amino acid sequence arg-ser-cys, which is part of the conserved sequence cys-cys-arg-ser-cys in the NS1 protein of D-2, D-4, JEV, YF, and WN. The individual primers which form the mixture in 5'-3 are shown in Fig. 45. In addition to the varied sequences derived from conserved region, each primer in each mixture also contains a constant region at the 5'-end which contains a sequence encoding sites for restriction enzymes, HindIII, MboI, and EcoRI.

The downstream primer, ssc5h20A, is derived from a nucleotide sequence in clone 5h, which contains HCV cDNA with sequences with overlap those in clones 14i and 11b. The sequence of ssc5h20A is

5' GTA ATA TGG TGA CAG AGT CA 3'.

An alternative primer, ssc5h34A, may also be used. This primer is derived from a sequence in clone 5h, and in addition contains nucleotides at the 5'-end which create a restriction enzyme site, thus facilitating cloning. The sequence of ssc5h34A is

5' GAT CTC TAG AGA AAT CAA TAT GGT GAC AGA GTC A 3'.

The PCR reaction, which was initially described by Saiki et al. (1986), is carried out essentially as described in Lee et al. (1988), except that the template for the cDNA is RNA isolated from HCV infected chimpanzee liver, as described in Section IV.C.2., or from viral particles isolated from HCV infected chimpanzee serum, as described in Section IV.A.1. In addition, the annealing conditions are less stringent in the first round of amplification (0.6M NaCl, and 25^{o}C), since the part of the primer which will anneal to the HCV sequence is only 9 nucleotides, and there could be mismatches. Moreover, if ssc5h34A is used, the additional sequences not derived from the HCV genome tend to destabilize the primer-template hybrid. After the first round of amplification, the annealing conditions can be more stringent (0.066M NaCl, and 32^{o}C-37^{o}C), since the amplified sequences now contain regions which are complementary to, or duplicates of the primers. In addition, the first 10 cycles of amplification are run with Klenow enzyme I, under appropriate PCR conditions for that enzyme. After the completion of these cycles, the samples are extracted, and run with Taq polymerase, according to kit directions, as furnished by Cetus/Perkin-Elmer.

After the amplification, the amplified HCV cDNA sequences are detected by hybridization using a probe derived from clone 5h. This probe is derived from sequences upstream of those used to derive the primer, and does not overlap the sequences of the clone 5h derived primers. The sequence of the probe is

5′ CCC AGC GGC GTA CGC GCT GGA CAC GGA GGT GGC CGC GTC
GTG TGG CGG TGT TGT TCT CGT CGG GTT GAT GGC GC 3′.

IV.N.1. Creation of HCV cDNA Library From Liver of a Chimpanzee with Infectious NANBH

An HCV cDNA library was created from liver from the chimpanzee from which the HCV cDNA library in Section IV.A.1. was created. The technique for creating the library was similar to that in Section IV.A.24, except for this different source of the RNA, and that a primer based on the sequence of HCV cDNA in clone 11b was used. The sequence of the primer was

5′ CTG GCT TGA AGA ATC 3′.

IV.N.2. Isolation and Nucleotide Sequence of Overlapping HCV cDNA in Clone k9-1 to cDNA in Clone 11b

Clone k9-1 was isolated from the HCV cDNA library created from the liver of an NANBH infected chimpanzee, as described in Section IV.A.25. The library was screened for clones which overlap the sequence in clone 11b, by using a clone which overlaps clone 11b at the 5′-terminus, clone 11e. The sequence of clone 11b is shown in Fig. 23. Positive clones were isolated with a frequency of 1 in 500,000. One isolated clone, k9-1, was subjected to further study. The overlapping nature of the HCV cDNA in clone k9-1 to the 5′-end of the HCV-cDNA sequence in Fig. 26 was confirmed by probing the clone with clone Alex46; this latter clone contains an HCV cDNA sequence of 30 base pairs which corresponds to those base pairs at the 5′terminus of the HCV cDNA in clone 14i, described supra.

The nucleotide sequence of the HCV cDNA isolated from clone k9-1 was determined using the techniques described supra. The sequence of the HCV cDNA in clone k9-1, the overlap with the HCV cDNA in Fig. 26, and the amino acids encoded therein are shown in Fig. 46.

The HCV cDNA sequence in clone k9-1 has been aligned with those of the clones described in Section IV.A.19. to create a composite HCV cDNA sequence, with the k9-1 sequence being placed upstream of the sequence shown in Fig. 32. The composite HCV cDNA which includes the k9-1 sequence, and the amino acids encoded therein, is shown in Fig. 47.

The sequence of the amino acids encoded in the 5'-region of HCV cDNA shown in Fig. 47 has been compared with the corresponding region of one of the strains of Dengue virus, described supra., with respect to the profile of regions of hydrophobicity and hydrophilicity. This comparison showed that the polypeptides from HCV and Dengue encoded in this region, which corresponds to the region encoding NS1 (or a portion thereof), have a similar hydrophobic/hydrophilic profile.

The information provided infra. allows the identification of HCV strains. The isolation and characterization of other HCV strains may be accomplished by isolating the nucleic acids from body components which contain viral particles, creating cDNA libraries using polynucleotide probes based on the HCV cDNA probes described infra., screening the libraries for clones containing HCV cDNA sequences described infra., and comparing the HCV cDNAs from the new isolates with the cDNAs described infra. The polypeptides encoded therein, or in the viral genome, may be monitored for immunological cross-reactivity utilizing the polypeptides and antibodies described supra. Strains which fit within the parameters of HCV, as described in the Definitions section, supra., are readily identifiable. Other methods for identifying HCV strains will be obvious to those of skill in the art, based upon the information provided herein.

Industrial Applicability

The invention, in the various manifestations disclosed herein, has many industrial uses, some of which are the following. The HCV cDNAs may be used for the design of probes for the detection of HCV nucleic acids in samples. The probes derived from the cDNAs may be used to detect HCV nucleic acids in, for example, chemical synthetic reactions. They may also be used in screening programs for anti-viral agents, to determine the effect of the agents in inhibiting viral replication in cell culture systems, and animal model systems. The HCV polynucleotide probes are also useful in detecting viral nucleic acids in humans, and thus, may serve as a basis for diagnosis of HCV infections in humans.

In addition to the above, the cDNAs provided herein provide information and a means for synthesizing polypeptides containing epitopes of HCV. These polypeptides are useful in detecting antibodies to HCV antigens. A series of immunoassays for HCV infection, based on recombinant polypeptides containing HCV epitopes are described herein, and will find commercial use in diagnosing HCV induced NANBH, in screening blood bank donors for HCV-caused infectious hepatitis, and also for detecting contaminated blood from infectious blood donors. The viral antigens will also have utility in monitoring the efficacy of anti-viral agents in animal model systems. In addition, the polypeptides derived from the HCV cDNAs disclosed herein will have utility as vaccines for treatment of HCV infections.

The polypeptides derived from the HCV cDNAs, besides the above stated uses, are also useful for raising anti-HCV antibodies. Thus, they may be used in anti-HCV vaccines. However, the antibodies produced as a result of immunization with the HCV polypeptides are also useful in detecting the presence of viral antigens in samples. Thus,

they may be used to assay the production of HCV polypeptides in chemical systems. The anti-HCV antibodies may also be used to monitor the efficacy of anti-viral agents in screening programs where these agents are tested in tissue culture systems. They may also be used for passive immunotherapy, and to diagnose HCV caused NANBH by allowing the detection of viral antigen(s) in both blood donors and recipients. Another important use for anti-HCV antibodies is in affinity chromatography for the purification of virus and viral polypeptides. The purified virus and viral polypeptide preparations may be used in vaccines. However, the purified virus may also be useful for the development of cell culture systems in which HCV replicates.

Cell culture systems containing HCV infected cells will have many uses. They can be used for the relatively large scale production of HCV, which is normally a low titer virus. These systems will also be useful for an elucidation of the molecular biology of the virus, and lead to the development of anti-viral agents. The cell culture systems will also be useful in screening for the efficacy of antiviral agents. In addition, HCV permissive cell culture systems are useful for the production of attenuated strains of HCV.

For convenience, the anti-HCV antibodies and HCV polypeptides, whether natural or recombinant, may be packaged into kits.

The method used for isolating HCV cDNA, which is comprised of preapring a cDNA library derived from infected tissue of an individual, in an expression vector, and selecting clones which produce the expression products which react immunologically with antibodies in antibody-containing body components from other infected individuals and not from non-infected individuals, may also be applicable to the isolation of cDNAs derived from other

heretofore uncharacterized disease-associated agents which are comprised of a genomic component. This, in turn, could lead to isolation and characterization of these agents, and to diagnostic reagents and vaccines for these other disease-associated agents.

CLAIMS

1. A purified HCV polynucleotide.

2. A recombinant HCV polynucleotide.

3. A recombinant polynucleotide comprising a sequence derived from an HCV genome or from HCV cDNA.

4. A recombinant polynucleotide encoding an epitope of HCV.

5. A recombinant vector containing the polynucleotide of claim 2, or claim 3, or claim 4.

6. A host cell transformed with the vector of claim 5.

7. A recombinant expression system comprising an open reading frame (ORF) of DNA derived from an HCV genome or from HCV cDNA, wherein the ORF is operably linked to a control sequence compatible with a desired host.

8. A cell transformed with the recombinant expression system of claim 7.

9. A polypeptide produced by the cell of claim 8.

10. Purified HCV.

11. A preparation of polypeptides from the HCV of claim 10.

12. A purified HCV polypeptide.

13. A purified polypeptide comprising an epitope which is immunologically identifiable with an epitope contained in HCV.

14. A recombinant HCV polypeptide.

15. A recombinant polypeptide comprised of a sequence derived from an HCV genome or from HCV cDNA.

16. A recombinant polypeptide comprised of an HCV epitope.

17. A fusion polypeptide comprised of an HCV polypeptide.

18. A monoclonal antibody directed against an HCV epitope.

19. A purified preparation of polyclonal antibodies directed against HCV.

20. A particle which is immunogenic against HCV infection comprising a non-HCV polypeptide having an amino acid sequence capable of forming a particle when said sequence is produced in a eukaryotic host, and an HCV epitope.

21. A polynucleotide probe for HCV.

22. A kit for analyzing samples for the presence of polynucleotides derived from HCV comprising a polynucleotide probe containing a nucleotide sequence from

HCV of about 8 or more nucleotides, in a suitable container.

23. A kit for analyzing samples for the presence of an HCV antigen comprising an antibody directed against the HCV antigen to be detected, in a suitable container.

24. A kit for analyzing samples for the presence of an antibodies directed against an HCV antigen comprising a polypeptide containing an HCV epitope present in the HCV antigen, in a suitable container.

25. A polypeptide comprised of an HCV epitope, attached to a solid substrate.

26. An antibody to an HCV epitope, attached to a solid substrate.

27. A method for producing a polypeptide containing an HCV epitope comprising incubating host cells transformed with an expression vector containing a sequence encoding a polypeptide containing an HCV epitope under conditions which allow expression of said polypeptide.

28. A polypeptide containing an HCV epitope produced by the method of claim 27.

29. A method for detecting HCV nucleic acids in a sample comprising:

(a) reacting nucleic acids of the sample with a probe for an HCV polynucleotide under conditions which allow the formation of a polynucleotide duplex between the probe and the HCV nucleic acid from the sample; and

(b) detecting a polynucleotide duplex which contains the probe.

30. An immunoassay for detecting an HCV antigen comprising:

(a) incubating a sample suspected of containing an HCV antigen with a probe antibody directed against the HCV antigen to be detected under conditions which allow the formation of an antigen-antibody complex; and

(b) detecting an antigen-antibody complex containing the probe antibody.

31. An immunoassay for detecting antibodies directed against an HCV antigen comprising:

(a) incubating a sample suspected of containing anti-HCV antibodies with a probe polypeptide which contains an epitope of the HCV, under conditions which allow the formation of an antibody-antigen complex; and

(b) detecting the antibody-antigen complex containing the probe antigen.

32. A vaccine for treatment of HCV infection comprising an immunogenic polypeptide containing an HCV epitope wherein the immunogenic polypeptide is present in a pharmacologically effective dose in a pharmaceutically acceptable excipient.

33. A vaccine for treatment of HCV infection comprising inactivated HCV in a pharmacologically effective dose in a pharmaceutically acceptable excipient.

34. A vaccine for treatment of HCV infection comprising attenuated HCV in a pharmacologically effective dose in a pharmaceutically acceptable excipient.

35. A tissue culture grown cell infected with HCV.

36. The HCV infected cell of claim 35, wherein the cell is of a human macrophage cell line, or is of a hepatocyte cell line, or is of a mosquito cell line, or is of a tick cell line, or is of a mouse macrophage cell line, or is an embryonic cell.

37. The HCV infected cell of claim 35, wherein the cell is of a cell line derived from liver of an HCV infected individual.

38. A method for producing antibodies to HCV comprising administering to an individual an isolated immunogenic polypeptide containing an HCV epitope in an amount sufficient to produce an immune response.

39. A method for producing antibodies to HCV comprising administering to an individual the polypeptide preparation of claim 11, wherein the preparation contains at least 1 immunogenic polypeptide, and the administering is of an amount sufficient to produce an immune response.

40. A method for isolating cDNA derived from the genome of an unidentified infectious agent, comprising:

(a) providing host cells transformed with expression vectors containing a cDNA library prepared from nucleic acids isolated from tissue infected with the agent and growing said host cells under conditions which allow expression of polypeptide(s) encoded in the cDNA;

(b) interacting the expression products of the cDNA with an antibody containing body component of an individual infected with said infectious agent under

conditions which allow an immunoreaction, and detecting antibody-antigen complexes formed as a result of the interacting;

(c) growing host cells which express polypeptides that form antibody-antigen complexes in step (b) under conditions which allow their growth as individual clones and isolating said clones;

(d) growing cells from the clones of (c) under conditions which allow expression of polypeptide(s) encoded within the cDNA, and interacting the expression products with antibody containing body components of individuals other than the individual in step (a) who are infected with the infectious agent and with control individuals uninfected with the agent, and detecting antibody-antigen complexes formed as a result of the interacting;

(e) growing host cells which express polypeptides that form antibody-antigen complexes with antibody containing body components of infected individuals and individuals suspected of being infected, and not with said components of control individuals, under conditions which allow their growth as individual clones and isolating said clones; and

(f) isolating the cDNA from the host cell clones of (e).

FIG. 1 Translation of DNA 5-1-1

```
      AlaSerCysLeuAsnCysSerAlaSerIleIleProAspArgGluValLeuTyrArgGlu
  1   GGCCTCCTGCTTGAACTGCTCGGCGAGCATCATACCTGACAGGGAAGTCCTCTACCGAGA
      CCGGAGGACGAACTTGACGAGCCGCTCGTAGTATGGACTGTCCCTTCAGGAGATGGCTCT

      PheAspGluMetGluGluCysSerGlnHisLeuProTyrIleGluGlnGlyMetMetLeu
 61   GTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCGTACATCGAGCAAGGGATGATGCT
      CAAGCTACTCTACCTTCTCACGAGAGTCGTGAATGGCATGTAGCTCGTTCCCTACTACGA

      AlaGluGlnPheLysGlnLysAlaLeuGlyLeu
121   CGCCGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCC
      GCGGCTCGTCAAGTTCGTCTTCCGGGAGCCGGAGG
```

FIG. 3 Translation of DNA 5-1-1,81,91&1-2

```
      GlyCysValValIleValGlyArgValValLeuSerGlyLysProAlaIleIleProAsp
  1   CTGGCTGCGTGGTCATAGTGGGCAGGGTCGTCTTGTCCGGGAAGCCGGCAATCATACCTG
      GACCGACGCACCAGTATCACCCGTCCCAGCAGAACAGGCCCTTCGGCCGTTAGTATGGAC

                        T
      ArgGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyr
 61   ACAGGGAAGTCCTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCGT
      TGTCCCTTCAGGAGATGGCTCTCAAGCTACTCTACCTTCTCACGAGAGTCGTGAATGGCA
                        A

      IleGluGlnGlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeuLeuGln
121   ACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCCTGC
      TGTAGCTCGTTCCCTACTACGAGCGGCTCGTCAAGTTCGTCTTCCGGGAGCCGGAGGACG

      ThrAlaSerArgGlnAlaGluValIleAlaProAlaValGlnThrAsnTrpGlnLysLeu
181   AGACCGCGTCCCGTCAGGCAGAGGTTATCGCCCCTGCTGTCCAGACCAACTGGCAAAAAC
      TCTGGCGCAGGGCAGTCCGTCTCCAATAGCGGGGACGACAGGTCTGGTTGACCGTTTTTG

      GluThrPheTrpAlaLysHisMetTrpAsnPheIleSerGlyIleGlnTyrLeuAlaGly
241   TCGAGACCTTCTGGGCGAAGCATATGTGGAACTTCATCAGTGGGATACAATACTTGGCGG
      AGCTCTGGAAGACCCGCTTCGTATACACCTTGAAGTAGTCACCCTATGTTATGAACCGCC

      LeuSerThrLeuProGlyAsnProAlaIleAlaSerLeuMetAlaPheThrAlaAlaVal
301   GCTTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACAGCTGCTG
      CGAACAGTTGCGACGGACCATTGGGGCGGTAACGAAGTAACTACCGAAAATGTCGACGAC

      ThrSerProLeuThrThrSerGln
361   TCACCAGCCCACTAACCACTAGCCAAA
      AGTGGTCGGGTGATTGGTGATCGGTTT
```

```
5-1-1   1    [ggcctcctgcttgaactgctcggcgagc]ATCATACCTGACAGGGAAG
81      1                       GTCCGGGAAGCCGGCAATCATACCTGACAGGGAAG
91      1    ctggctgcgtGGTCATAGTGGCAGGGTCGTCTTGTCCGGGAAGCCGGCAATCATACCTGACAGGGAAG
1-2     1              GGTCATAGTGGCAGGGTCGTCTTGTCCGGGAAGCCGGCAATCATACCTGACAGGGAAG

5-1-1   48   TCCTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCGTACATCGAGCAAGGGATGATGC
81      36   TCCTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCGTACATCGAGCAAGGGATGATGC
91      70   TCCTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCGTACATCGAGCAAGGGATGATGC
1-2     60   TCCTCTAtCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCGTACATCGAGCAAGGGATGATGC

5-1-1   120  TCGCCGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCC
81      108  TCGCCGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCCTGCAGACCGCGTCCCGTCAGGCAGAGGTTATCGCCC
91      142  TCGCCGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCCTGCAGACCGCGTCCCGTCAGGCAGAGGTTATCGCCC
1-2     132  TCGCCGAGCAGTTCAAGCAGAAGGCCCTCGGCC

81      180  CTGCTGTCCAGACCAACTGGCAAAAACTCGAGACCTTCTGGGCGAAGCATATGTGGAACTTCATCAGTGGGA
91      214  CTGCTGTCCAGACCAACTGGCAAAAACTCGAGACCTTCTGGGCGAAGCATATGTGGAACTTCATCAGTGGGA

81      252  TACAATACTTGGCGGGCTTGTCAACGCTGCCTGGtaaccccgccattgcttcattgatggcttttacagctg
91      286  TACAATACTTGGCGGGCTTGTCAACGCTGCCTGG

81      324  ctgtcaccagcccactaaccactagccaaa
```

FIG. 2

FIG. 4 Translation of DNA 81

```
     SerGlyLysProAlaIleIleProAspArgGluValLeuTyrArgGluPheAspGluMet
1    GTCCGGGAAGCCGGCAATCATACCTGACAGGGAAGTCCTCTACCGAGAGTTCGATGAGAT
     CAGGCCCTTCGGCCGTTAGTATGGACTGTCCCTTCAGGAGATGGCTCTCAAGCTACTCTA

     GluGluCysSerGlnHisLeuProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPhe
61   GGAAGAGTGCTCTCAGCACTTACCGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTT
     CCTTCTCACGAGAGTCGTGAATGGCATGTAGCTCGTTCCCTACTACGAGCGGCTCGTCAA

     LysGlnLysAlaLeuGlyLeuLeuGlnThrAlaSerArgGlnAlaGluValIleAlaPro
121  CAAGCAGAAGGCCCTCGGCCTCCTGCAGACCGCGTCCCGTCAGGCAGAGGTTATCGCCCC
     GTTCGTCTTCCGGGAGCCGGAGGACGTCTGGCGCAGGGCAGTCCGTCTCCAATAGCGGGG

     AlaValGlnThrAsnTrpGlnLysLeuGluThrPheTrpAlaLysHisMetTrpAsnPhe
181  TGCTGTCCAGACCAACTGGCAAAAACTCGAGACCTTCTGGGCGAAGCATATGTGGAACTT
     ACGACAGGTCTGGTTGACCGTTTTTGAGCTCTGGAAGACCCGCTTCGTATACACCTTGAA

     IleSerGlyIleGlnTyrLeuAlaGlyLeuSerThrLeuProGlyAsnProAlaIleAla
241  CATCAGTGGGATACAATACTTGGCGGGCTTGTCAACGCTGCCTGGTAACCCCGCCATTGC
     GTAGTCACCCTATGTTATGAACCGCCCGAACAGTTGCGACGGACCATTGGGGCGGTAACG

     SerLeuMetAlaPheThrAlaAlaValThrSerProLeuThrThrSerGln
301  TTCATTGATGGCTTTTACAGCTGCTGTCACCAGCCCACTAACCACTAGCCAAA
     AAGTAACTACCGAAAATGTCGACGACAGTGGTCGGGTGATTGGTGATCGGTTT
```

FIG. 5 Translation of DNA 36

```
     AspAlaHisPheLeuSerGlnThrLysGlnSerGlyGluAsnLeuProTyrLeuValAla
1    GATGCCCACTTTCTATCCCAGACAAAGCAGAGTGGGGAGAACCTTCCTTACCTGGTAGCG
     CTACGGGTGAAAGATAGGGTCTGTTTCGTCTCACCCCTCTTGGAAGGAATGGACCATCGC

     TyrGlnAlaThrValCysAlaArgAlaGlnAlaProProProSerTrpAspGlnMetTrp
61   TACCAAGCCACCGTGTGCGCTAGGGCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGG
     ATGGTTCGGTGGCACACGCGATCCCGAGTTCGGGGAGGGGGTAGCACCCTGGTCTACACC

     LysCysLeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu
121  AAGTGTTTGATTCGCCTCAAGCCCACCCTCCATGGGCCAACACCCCTGCTATACAGACTG
     TTCACAAACTAAGCGGAGTTCGGGTGGGAGGTACCCGGTTGTGGGGACGATATGTCTGAC

     GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIleMetThrCys
181  GGCGCTGTTCAGAATGAAATCACCCTGACGCACCCAGTCACCAAATACATCATGACATGC
     CCGCGACAAGTCTTACTTTAGTGGGACTGCGTGGGTCAGTGGTTTATGTAGTACTGTACG

     MetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyGlyValLeuAla
241  ATGTCGGCCGACCTGGAGGTCGTCACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCT
     TACAGCCGGCTGGACCTCCAGCAGTGCTCGTGGACCCACGAGCAACCGCCGCAGGACCGA

     AlaLeuAlaAlaTyrCysLeuSerThrGlyCysValValIleValGlyArgValValLeu
301  GCTTTGGCCGCGTATTGCCTGTCAACAGGCTGCGTGGTCATAGTGGGCAGGGTCGTCTTG
     CGAAACCGGCGCATAACGGACAGTTGTCCGACGCACCAGTATCACCCGTCCCAGCAGAAC

     -----------Overlap with 81-------------------
     SerGlyLysProAlaIleIleProAspArgGluValLeuTyrArg
361  TCCGGGAAGCCGGCAATCATACCTGACAGGGAAGTCCTCTACCGAG
     AGGCCCTTCGGCCGTTAGTATGGACTGTCCCTTCAGGAGATGGCTC
```

FIG. 6 Combined ORF of DNAs 36 & 81

```
     AspAlaHisPheLeuSerGlnThrLysGlnSerGlyGluAsnLeuProTyrLeuValAla
  1  GATGCCCACTTTCTATCCCAGACAAAGCAGAGTGGGGAGAACCTTCCTTACCTGGTAGCG
     CTACGGGTGAAAGATAGGGTCTGTTTCGTCTCACCCCTCTTGGAAGGAATGGACCATCGC

     TyrGlnAlaThrValCysAlaArgAlaGlnAlaProProProSerTrpAspGlnMetTrp
 61  TACCAAGCCACCGTGTGCGCTAGGGCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGG
     ATGGTTCGGTGGCACACGCGATCCCGAGTTCGGGGAGGGGGTAGCACCCTGGTCTACACC

     LysCysLeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu
121  AAGTGTTTGATTCGCCTCAAGCCCACCCTCCATGGGCCAACACCCCTGCTATACAGACTG
     TTCACAAACTAAGCGGAGTTCGGGTGGGAGGTACCCGGTTGTGGGGACGATATGTCTGAC

     GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIleMetThrCys
181  GGCGCTGTTCAGAATGAAATCACCCTGACGCACCCAGTCACCAAATACATCATGACATGC
     CCGCGACAAGTCTTACTTTAGTGGGACTGCGTGGGTCAGTGGTTTATGTAGTACTGTACG

     MetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyGlyValLeuAla
241  ATGTCGGCCGACCTGGAGGTCGTCACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCT
     TACAGCCGGCTGGACCTCCAGCAGTGCTCGTGGACCCACGAGCAACCGCCGCAGGACCGA

     AlaLeuAlaAlaTyrCysLeuSerThrGlyCysValValIleValGlyArgValValLeu
301  GCTTTGGCCGCGTATTGCCTGTCAACAGGCTGCGTGGTCATAGTGGGCAGGGTCGTCTTG
     CGAAACCGGCGCATAACGGACAGTTGTCCGACGCACCAGTATCACCCGTCCCAGCAGAAC

     SerGlyLysProAlaIleIleProAspArgGluValLeuTyrArgGluPheAspGluMet
361  TCCGGGAAGCCGGCAATCATACCTGACAGGGAAGTCCTCTACCGAGAGTTCGATGAGATG
     AGGCCCTTCGGCCGTTAGTATGGACTGTCCCTTCAGGAGATGGCTCTCAAGCTACTCTAC

     GluGluCysSerGlnHisLeuProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPhe
421  GAAGAGTGCTCTCAGCACTTACCGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTC
     CTTCTCACGAGAGTCGTGAATGGCATGTAGCTCGTTCCCTACTACGAGCGGCTCGTCAAG

     LysGlnLysAlaLeuGlyLeuLeuGlnThrAlaSerArgGlnAlaGluValIleAlaPro
481  AAGCAGAAGGCCCTCGGCCTCCTGCAGACCGCGTCCCGTCAGGCAGAGGTTATCGCCCCT
     TTCGTCTTCCGGGAGCCGGAGGACGTCTGGCGCAGGGCAGTCCGTCTCCAATAGCGGGGA

     AlaValGlnThrAsnTrpGlnLysLeuGluThrPheTrpAlaLysHisMetTrpAsnPhe
541  GCTGTCCAGACCAACTGGCAAAAACTCGAGACCTTCTGGGCGAAGCATATGTGGAACTTC
     CGACAGGTCTGGTTGACCGTTTTTGAGCTCTGGAAGACCCGCTTCGTATACACCTTGAAG

     IleSerGlyIleGlnTyrLeuAlaGlyLeuSerThrLeuProGlyAsnProAlaIleAla
601  ATCAGTGGGATACAATACTTGGCGGGCTTGTCAACGCTGCCTGGTAACCCCGCCATTGCT
     TAGTCACCCTATGTTATGAACCGCCCGAACAGTTGCGACGGACCATTGGGGCGGTAACGA

     SerLeuMetAlaPheThrAlaAlaValThrSerProLeuThrThrSerGln
661  TCATTGATGGCTTTTACAGCTGCTGTCACCAGCCCACTAACCACTAGCCAAA
     AGTAACTACCGAAAATGTCGACGACAGTGGTCGGGTGATTGGTGATCGGTTT
```

FIG. 7 Translation of DNA 32

```
      -------Overlap with 81---------------------
        PheThrAlaAlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsnIleLeu
  1   CTTTTACAGCTGCTGTCACCAGCCCACTAACCACTAGCCAAACCCTCCTCTTCAACATAT
      GAAAATGTCGACGACAGTGGTCGGGTGATTGGTGATCGGTTTGGGAGGAGAAGTTGTATA

        GlyGlyTrpValAlaAlaGlnLeuAlaAlaProGlyAlaAlaThrAlaPheValGlyAla
 61   TGGGGGGGTGGGTGGCTGCCCAGCTCGCCGCCCCCGGTGCCGCTACTGCCTTTGTGGGCG
      ACCCCCCCACCCACCGACGGGTCGAGCGGCGGGGGCCACGGCGATGACGGAAACACCCGC

        GlyLeuAlaGlyAlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAspIleLeu
121   CTGGCTTAGCTGGCGCCGCCATCGGCAGTGTTGGACTGGGGAAGGTCCTCATAGACATCC
      GACCGAATCGACCGCGGCGGTAGCCGTCACAACCTGACCCCTTCCAGGAGTATCTGTAGG

        AlaGlyTyrGlyAlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSerGlyGlu
181   TTGCAGGGTATGGCGCGGGCGTGGCGGGAGCTCTTGTGGCATTCAAGATCATGAGCGGTG
      AACGTCCCATACCGCGCCCGCACCGCCCTCGAGAACACCGTAAGTTCTAGTACTCGCCAC

        ValProSerThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGlyAlaLeu
241   AGGTCCCCTCCACGGAGGACCTGGTCAATCTACTGCCCGCCATCCTCTCGCCCGGAGCCC
      TCCAGGGGAGGTGCCTCCTGGACCAGTTAGATGACGGGCGGTAGGAGAGCGGGCCTCGGG

        ValValGlyValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGluGlyAla
301   TCGTAGTCGGCGTGGTCTGTGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAGGGGG
      AGCATCAGCCGCACCAGACACGTCGTTATGACGCGGCCGTGCAACCGGGCCCGCTCCCCC

        ValGlnTrpMetAsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSer
361   CAGTGCAGTGGATGAACCGGCTGATAGCCTTCGCCTCCCGGGGGAACCATGTTTCCCC
      GTCACGTCACCTACTTGGCCGACTATCGGAAGCGGAGGGCCCCCTTGGTACAAAGGGG
```

FIG. 8 Translation of DNA 35

```
    SerIleGluThrIleThrLeuProGlnAspAlaValSerArgThrGlnArgArgGlyArg
1   TCCATTGAGACAATCACGCTCCCCCAGGATGCTGTCTCCCGCACTCAACGTCGGGGCAGG
    AGGTAACTCTGTTAGTGCGAGGGGGTCCTACGACAGAGGGCGTGAGTTGCAGCCCCGTCC

    ThrGlyArgGlyLysProGlyIleTyrArgPheValAlaProGlyGluArgProSerGly
61  ACTGGCAGGGGGAAGCCAGGCATCTACAGATTTGTGGCACCGGGGGAGCGCCCCTCCGGC
    TGACCGTCCCCCTTCGGTCCGTAGATGTCTAAACACCGTGGCCCCCTCGCGGGGAGGCCG

    MetPheAspSerSerValLeuCysGluCysTyrAspAlaGlyCysAlaTrpTyrGluLeu
121 ATGTTCGACTCGTCCGTCCTCTGTGAGTGCTATGACGCAGGCTGTGCTTGGTATGAGCTC
    TACAAGCTGAGCAGGCAGGAGACACTCACGATACTGCGTCCGACACGAACCATACTCGAG

    ThrProAlaGluThrThrValArgLeuArgAlaTyrMetAsnThrProGlyLeuProVal
181 ACGCCCGCCGAGACTACAGTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTG
    TGCGGGCGGCTCTGATGTCAATCCGATGCTCGCATGTACTTGTGGGGCCCCGAAGGGCAC

                                                          ------
    CysGlnAspHisLeuGluPheTrpGluGlyValPheThrGlyLeuThrHisIleAspAla
241 TGCCAGGACCATCTTGAATTTTGGGAGGGCGTCTTTACAGGCCTCACTCATATAGATGCC
    ACGGTCCTGGTAGAACTTAAAACCCTCCCGCAGAAATGTCCGGAGTGAGTATATCTACGG

    ------------------------------------------------------------
    HisPheLeuSerGlnThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGln
301 CACTTTCTATCCCAGACAAAGCAGAGTGGGGAGAACCTTCCTTACCTGGTAGCGTACCAA
    GTGAAAGATAGGGTCTGTTTCGTCTCACCCCTCTTGGAAGGAATGGACCATCGCATGGTT

    ----------------Overlap with 36-----------------------------
    AlaThrValCysAlaArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLysCys
361 GCCACCGTGTGCGCTAGGGCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGT
    CGGTGGCACACGCGATCCCGAGTTCGGGGAGGGGGTAGCACCCTGGTCTACACCTTCACA

    ------------------------------------------------------------
    LeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAla
421 TTGATTCGCCTCAAGCCCACCCTCCATGGCCAACACCCCTGCTATACAGACTGGGCGCT
    AACTAAGCGGAGTTCGGGTGGGAGGTACCCGGTTGTGGGGACGATATGTCTGACCCGCGA
```

FIG. 9-1 Combined ORF of DNAs 35,36,81 & 32

```
    SerIleGluThrIleThrLeuProGlnAspAlaValSerArgThrGlnArgArgGlyArg
  1 TCCATTGAGACAATCACGCTCCCCCAGGATGCTGTCTCCCGCACTCAACGTCGGGGCAGG
    AGGTAACTCTGTTAGTGCGAGGGGGTCCTACGACAGAGGGCGTGAGTTGCAGCCCCGTCC

    ThrGlyArgGlyLysProGlyIleTyrArgPheValAlaProGlyGluArgProSerGly
 61 ACTGGCAGGGGGAAGCCAGGCATCTACAGATTTGTGGCACCGGGGGAGCGCCCCTCCGGC
    TGACCGTCCCCCTTCGGTCCGTAGATGTCTAAACACCGTGGCCCCCTCGCGGGGAGGCCG

    MetPheAspSerSerValLeuCysGluCysTyrAspAlaGlyCysAlaTrpTyrGluLeu
121 ATGTTCGACTCGTCCGTCCTCTGTGAGTGCTATGACGCAGGCTGTGCTTGGTATGAGCTC
    TACAAGCTGAGCAGGCAGGAGACACTCACGATACTGCGTCCGACACGAACCATACTCGAG

    ThrProAlaGluThrThrValArgLeuArgAlaTyrMetAsnThrProGlyLeuProVal
181 ACGCCCGCCGAGACTACAGTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTG
    TGCGGGCGGCTCTGATGTCAATCCGATGCTCGCATGTACTTGTGGGGCCCCGAAGGGCAC

    CysGlnAspHisLeuGluPheTrpGluGlyValPheThrGlyLeuThrHisIleAspAla
241 TGCCAGGACCATCTTGAATTTTGGGAGGGCGTCTTTACAGGCCTCACTCATATAGATGCC
    ACGGTCCTGGTAGAACTTAAAACCCTCCCGCAGAAATGTCCGGAGTGAGTATATCTACGG

    HisPheLeuSerGlnThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGln
301 CACTTTCTATCCCAGACAAAGCAGAGTGGGGAGAACCTTCCTTACCTGGTAGCGTACCAA
    GTGAAAGATAGGGTCTGTTTCGTCTCACCCCTCTTGGAAGGAATGGACCATCGCATGGTT

    AlaThrValCysAlaArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLysCys
361 GCCACCGTGTGCGCTAGGGCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGT
    CGGTGGCACACGCGATCCCGAGTTCGGGGAGGGGGTAGCACCCTGGTCTACACCTTCACA

    LeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAla
421 TTGATTCGCCTCAAGCCCACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCT
    AACTAAGCGGAGTTCGGGTGGGAGGTACCCGGTTGTGGGGACGATATGTCTGACCCGCGA

    ValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIleMetThrCysMetSer
481 GTTCAGAATGAAATCACCCTGACGCACCCAGTCACCAAATACATCATGACATGCATGTCG
    CAAGTCTTACTTTAGTGGGACTGCGTGGGTCAGTGGTTTATGTAGTACTGTACGTACAGC

    AlaAspLeuGluValValThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeu
541 GCCGACCTGGAGGTCGTCACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTTTG
    CGGCTGGACCTCCAGCAGTGCTCGTGGACCCACGAGCAACCGCCGCAGGACCGACGAAAC

    AlaAlaTyrCysLeuSerThrGlyCysValValIleValGlyArgValValLeuSerGly
601 GCCGCGTATTGCCTGTCAACAGGCTGCGTGGTCATAGTGGGCAGGGTCGTCTTGTCCGGG
    CGGCGCATAACGGACAGTTGTCCGACGCACCAGTATCACCCGTCCCAGCAGAACAGGCCC

    LysProAlaIleIleProAspArgGluValLeuTyrArgGluPheAspGluMetGluGlu
661 AAGCCGGCAATCATACCTGACAGGGAAGTCCTCTACCGAGAGTTCGATGAGATGGAAGAG
    TTCGGCCGTTAGTATGGACTGTCCCTTCAGGAGATGGCTCTCAAGCTACTCTACCTTCTC

    CysSerGlnHisLeuProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPheLysGln
721 TGCTCTCAGCACTTACCGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAG
    ACGAGAGTCGTGAATGGCATGTAGCTCGTTCCCTACTACGAGCGGCTCGTCAAGTTCGTC

    LysAlaLeuGlyLeuLeuGlnThrAlaSerArgGlnAlaGluValIleAlaProAlaVal
781 AAGGCCCTCGGCCTCCTGCAGACCGCGTCCCGTCAGGCAGAGGTTATCGCCCCTGCTGTC
    TTCCGGGAGCCGGAGGACGTCTGGCGCAGGGCAGTCCGTCTCCAATAGCGGGGACGACAG
```

```
     GlnThrAsnTrpGlnLysLeuGluThrPheTrpAlaLysHisMetTrpAsnPheIleSer
841  CAGACCAACTGGCAAAAACTCGAGACCTTCTGGGCGAAGCATATGTGGAACTTCATCAGT
     GTCTGGTTGACCGTTTTTGAGCTCTGGAAGACCCGCTTCGTATACACCTTGAAGTAGTCA

     GlyIleGlnTyrLeuAlaGlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeu
901  GGGATACAATACTTGGCGGGCTTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTG
     CCCTATGTTATGAACCGCCCGAACAGTTGCGACGGACCATTGGGGCGGTAACGAAGTAAC

     MetAlaPheThrAlaAlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsn
961  ATGGCTTTTACAGCTGCTGTCACCAGCCCACTAACCACTAGCCAAACCCTCCTCTTCAAC
     TACCGAAAATGTCGACGACAGTGGTCGGGTGATTGGTGATCGGTTTGGGAGGAGAAGTTG

     IleLeuGlyGlyTrpValAlaAlaGlnLeuAlaAlaProGlyAlaAlaThrAlaPheVal
1021 ATATTGGGGGGGTGGGTGGCTGCCCAGCTCGCCGCCCCCGGTGCCGCTACTGCCTTTGTG
     TATAACCCCCCCACCCACCGACGGGTCGAGCGGCGGGGGCCACGGCGATGACGGAAACAC

     GlyAlaGlyLeuAlaGlyAlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAsp
1081 GGCGCTGGCTTAGCTGGCGCCGCCATCGGCAGTGTTGGACTGGGGAAGGTCCTCATAGAC
     CCGCGACCGAATCGACCGCGGCGGTAGCCGTCACAACCTGACCCCTTCCAGGAGTATCTG

     IleLeuAlaGlyTyrGlyAlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSer
1141 ATCCTTGCAGGGTATGGCGCGGGCGTGGCGGGAGCTCTTGTGGCATTCAAGATCATGAGC
     TAGGAACGTCCCATACCGCGCCCGCACCGCCCTCGAGAACACCGTAAGTTCTAGTACTCG

     GlyGluValProSerThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGly
1201 GGTGAGGTCCCCTCCACGGAGGACCTGGTCAATCTACTGCCCGCCATCCTCTCGCCCGGA
     CCACTCCAGGGGAGGTGCCTCCTGGACCAGTTAGATGACGGGCGGTAGGAGAGCGGGCCT

     AlaLeuValValGlyValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGlu
1261 GCCCTCGTAGTCGGCGTGGTCTGTGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAG
     CGGGAGCATCAGCCGCACCAGACACGTCGTTATGACGCGGCCGTGCAACCGGGCCCGCTC

     GlyAlaValGlnTrpMetAsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSer
1321 GGGGCAGTGCAGTGGATGAACCGGCTGATAGCCTTCGCCTCCCGGGGGAACCATGTTTCCCC
     CCCCGTCACGTCACCTACTTGGCCGACTATCGGAAGCGGAGGGCCCCCTTGGTACAAAGGGG
```

FIG. 9-2

FIG. 10 Translation of DNA 37b

```
     LeuAlaAlaLysLeuValAlaLeuGlyIleAsnAlaValAlaTyrTyrArgGlyLeuAsp
  1  CTCGCCGCAAAGCTGGTCGCATTGGGCATCAATGCCGTGGCCTACTACCGCGGTCTTGAC
     GAGCGGCGTTTCGACCAGCGTAACCCGTAGTTACGGCACCGGATGATGGCGCCAGAACTG

     ValSerValIleProThrSerGlyAspValValValValAlaThrAspAlaLeuMetThr
 61  GTGTCCGTCATCCCGACCAGCGGCGATGTTGTCGTCGTGGCAACCGATGCCCTCATGACC
     CACAGGCAGTAGGGCTGGTCGCCGCTACAACAGCAGCACCGTTGGCTACGGGAGTACTGG

     GlyTyrThrGlyAspPheAspSerValIleAspTyrAsnThrCysValThrGlnThrVal
121  GGCTATACCGGCGACTTCGACTCGGTGATAGACTACAATACGTGTGTCACCCAGACAGTC
     CCGATATGGCCGCTGAAGCTGAGCCACTATCTGATGTTATGCACACAGTGGGTCTGTCAG

                                              ----------------------Overlap with
     AspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuProGlnAspAlaVal
181  GATTTCAGCCTTGACCCTACCTTCACCATTGAGACAATCACGCTCCCCCAGGATGCTGTC
     CTAAAGTCGGAACTGGGATGGAAGTGGTAACTCTGTTAGTGCGAGGGGGTCCTACGACAG

     clone 35--------------------
     SerArgThrGlnArgArgGlyArgThr
241  TCCCGCACTCAACGTCGGGGCAGGACTG
     AGGGCGTGAGTTGCAGCCCCGTCCTGAC
```

FIG. 11 Translation of DNA 33b

```
     -------Overlap with 32---------------------------
      MetAsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSerProThrHisTyrVal
  1  GATGAACCGGCTGATAGCCTTCGCCTCCCGGGGGAACCATGTTTCCCCCACGCACTACGT
     CTACTTGGCCGACTATCGGAAGCGGAGGGCCCCCTTGGTACAAAGGGGGTGCGTGATGCA

      ProGluSerAspAlaAlaAlaArgValThrAlaIleLeuSerSerLeuThrValThrGln
 61  GCCGGAGAGCGATGCAGCTGCCCGCGTCACTGCCATACTCAGCAGCCTCACTGTAACCCA
     CGGCCTCTCGCTACGTCGACGGGCGCAGTGACGGTATGAGTCGTCGGAGTGACATTGGGT

      LeuLeuArgArgLeuHisGlnTrpIleSerSerGluCysThrThrProCysSerGlySer
121  GCTCCTGAGGCGACTGCACCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGGTTC
     CGAGGACTCCGCTGACGTGGTCACCTATTCGAGCCTCACATGGTGAGGTACGAGGCCAAG

      TrpLeuArgAspIleTrpAspTrpIleCysGluValLeuSerAspPheLysThrTrpLeu
181  CTGGCTAAGGGACATCTGGGACTGGATATGCGAGGTGTTGAGCGACTTTAAGACCTGGCT
     GACCGATTCCCTGTAGACCCTGACCTATACGCTCCACAACTCGCTGAAATTCTGGACCGA

      LysAlaLysLeuMetProGlnLeuProGlyIleProPheValSerCysGlnArgGlyTyr
241  AAAAGCTAAGCTCATGCCACAGCTGCCTGGGATCCCCTTTGTGTCCTGCCAGCGCGGGTA
     TTTTCGATTCGAGTACGGTGTCGACGGACCCTAGGGGAAACACAGGACGGTCGCGCCCAT

      LysGlyValTrpArgVal
301  TAAGGGGGTCTGGCGAGTG
     ATTCCCCCAGACCGCTCAC
```

FIG. 12 Translation of DNA 40b

```
     AlaTyrMetSerLysAlaHisGlyIleAspProAsnIleArgThrGlyValArgThrIle
  1  GGCTTACATGTCCAAGGCTCATGGGATCGATCCTAACATCAGGACCGGGGTGAGAACAAT
     CCGAATGTACAGGTTCCGAGTACCCTAGCTAGGATTGTAGTCCTGGCCCCACTCTTGTTA

     ThrThrGlySerProIleThrTyrSerThrTyrGlyLysPheLeuAlaAspGlyGlyCys
 61  TACCACTGGCAGCCCCATCACGTACTCCACCTACGGCAAGTTCCTTGCCGACGGCGGGTG
     ATGGTGACCGTCGGGGTAGTGCATGAGGTGGATGCCGTTCAAGGAACGGCTGCCGCCCAC

     SerGlyGlyAlaTyrAspIleIleIleCysAspGluCysHisSerThrAspAlaThrSer
121  CTCGGGGGCGCTTATGACATAATAATTTGTGACGAGTGCCACTCCACGGATGCCACATC
     GAGCCCCCGCGAATACTGTATTATTAAACACTGCTCACGGTGAGGTGCCTACGGTGTAG

     IleLeuGlyIleGlyThrValLeuAspGlnAlaGluThrAlaGlyAlaArgLeuValVal
181  CATCTTGGGCATCGGCACTGTCCTTGACCAAGCAGAGACTGCGGGGGCGAGACTGGTTGT
     GTAGAACCCGTAGCCGTGACAGGAACTGGTTCGTCTCTGACGCCCCCGCTCTGACCAACA

     LeuAlaThrAlaThrProProGlySerValThrValProHisProAsnIleGluGluVal
241  GCTCGCCACCGCCACCCCTCCGGGCTCCGTCACTGTGCCCCATCCCAACATCGAGGAGGT
     CGAGCGGTGGCGGTGGGGAGGCCCGAGGCAGTGACACGGGGTAGGGTTGTAGCTCCTCCA

     AlaLeuSerThrThrGlyGluIleProPheTyrGlyLysAlaIleProLeuGluValIle
301  TGCTCTGTCCACCACCGGAGAGATCCCTTTTTACGGCAAGGCTATCCCCCTCGAAGTAAT
     ACGAGACAGGTGGTGGCCTCTCTAGGGAAAAATGCCGTTCCGATAGGGGGAGCTTCATTA
                                                             --------
     LysGlyGlyArgHisLeuIlePheCysHisSerLysLysLysCysAspGluLeuAlaAla
361  CAAGGGGGGAGACATCTCATCTTCTGTCATTCAAAGAAGAAGTGCGACGAACTCGCCGC
     GTTCCCCCCCTCTGTAGAGTAGAAGACAGTAAGTTTCTTCTTCACGCTGCTTGAGCGGCG

     ----------------------Overlap with 37b---------------------
     LysLeuValAlaLeuGlyIleAsnAlaValAlaTyrTyrArgGlyLeuAspValSerVal
421  AAAGCTGGTCGCATTGGGCATCAATGCCGTGGCCTACTACCGCGGTCTTGACGTGTCCGT
     TTTCGACCAGCGTAACCCGTAGTTACGGCACCGGATGATGGCGCCAGAACTGCACAGGCA

     ---------
     IleProThr
481  CATCCCGACCAG
     GTAGGGCTGGTC
```

FIG. 13 Translation of DNA 25c

```
     ------------------------------------------------------------
       CysSerLeuThrValThrGlnLeuLeuArgArgLeuHisGlnTrpIleSerSerGluCys
  1  ACTGCAGCCTCACTGTAACCCAGCTCCTGAGGCGACTGCACCAGTGGATAAGCTCGGAGT
     TGACGTCGGAGTGACATTGGGTCGAGGACTCCGCTGACGTGGTCACCTATTCGAGCCTCA

     ------------------------------------------------------------
       ThrThrProCysSerGlySerTrpLeuArgAspIleTrpAspTrpIleCysGluValLeu
 61  GTACCACTCCATGCTCCGGTTCCTGGCTAAGGGACATCTGGGACTGGATATGCGAGGTGT
     CATGGTGAGGTACGAGGCCAAGGACCGATTCCCTGTAGACCCTGACCTATACGCTCCACA

     ---------------Overlap with 33b-----------------------------
       SerAspPheLysThrTrpLeuLysAlaLysLeuMetProGlnLeuProGlyIleProPhe
121  TGAGCGACTTTAAGACCTGGCTAAAAGCTAAGCTCATGCCACAGCTGCCTGGGATCCCCT
     ACTCGCTGAAATTCTGGACCGATTTTCGATTCGAGTACGGTGTCGACGGACCCTAGGGGA

     ----------------------------------------
       ValSerCysGlnArgGlyTyrLysGlyValTrpArgGlyAspGlyIleMetHisThrArg
181  TTGTGTCCTGCCAGCGCGGGTATAAGGGGGTCTGGCGAGGGGACGGCATCATGCACACTC
     AACACAGGACGGTCGCGCCCATATTCCCCCAGACCGCTCCCCTGCCGTAGTACGTGTGAG

       CysHisCysGlyAlaGluIleThrGlyHisValLysAsnGlyThrMetArgIleValGly
241  GCTGCCACTGTGGAGCTGAGATCACTGGACATGTCAAAAACGGGACGATGAGGATCGTCG
     CGACGGTGACACCTCGACTCTAGTGACCTGTACAGTTTTTGCCCTGCTACTCCTAGCAGC

       ProArgThrCysArgAsnMetTrpSerGlyThrPheProIleAsnAlaTyrThrThrGly
301  GTCCTAGGACCTGCAGGAACATGTGGAGTGGGACCTTCCCCATTAATGCCTACACCACGG
     CAGGATCCTGGACGTCCTTGTACACCTCACCCTGGAAGGGGTAATTACGGATGTGGTGCC

       ProCysThrProLeuProAlaProAsnTyrThrPheAlaLeuTrpArgValSerAlaGlu
361  GCCCCTGTACCCCCCTTCCTGCGCCGAACTACACGTTCGCGCTATGGAGGGTGTCTGCAG
     CGGGGACATGGGGGGAAGGACGCGGCTTGATGTGCAAGCGCGATACCTCCCACAGACGTC

       GluTyrValGluIleArgGlnValGlyAspPheHisTyrValThrGlyMetThrThrAsp
421  AGGAATATGTGGAGATAAGGCAGGTGGGGGACTTCCACTACGTGACGGGTATGACTACTG
     TCCTTATACACCTCTATTCCGTCCACCCCCTGAAGGTGATGCACTGCCCATACTGATGAC

       AsnLeuLysCysProCysGlnValProSerProGluPhePheThrGlu
481  ACAATCTCAAATGCCCGTGCCAGGTCCCATCGCCCGAATTTTTCACAGAAT
     TGTTAGAGTTTACGGCACGGTCCAGGGTAGCGGGCTTAAAAAGTGTCTTA
```

FIG. 14-1 Combined ORF of DNAs 40b/37b/35/36/81/32/33b/25c

```
    AlaTyrMetSerLysAlaHisGlyIleAspProAsnIleArgThrGlyValArgThrIle
1   TGCTTACATGTCCAAGGCTCATGGGATCGATCCTAACATCAGGACCGGGGTGAGAACAAT
    ACGAATGTACAGGTTCCGAGTACCCTAGCTAGGATTGTAGTCCTGGCCCCACTCTTGTTA

    ThrThrGlySerProIleThrTyrSerThrTyrGlyLysPheLeuAlaAspGlyGlyCys
61  TACCACTGGCAGCCCCATCACGTACTCCACCTACGGCAAGTTCCTTGCCGACGGCGGGTG
    ATGGTGACCGTCGGGGTAGTGCATGAGGTGGATGCCGTTCAAGGAACGGCTGCCGCCCAC

    SerGlyGlyAlaTyrAspIleIleIleCysAspGluCysHisSerThrAspAlaThrSer
121 CTCGGGGGGCGCTTATGACATAATAATTTGTGACGAGTGCCACTCCACGGATGCCACATC
    GAGCCCCCCGCGAATACTGTATTATTAAACACTGCTCACGGTGAGGTGCCTACGGTGTAG

    IleLeuGlyIleGlyThrValLeuAspGlnAlaGluThrAlaGlyAlaArgLeuValVal
181 CATCTTGGGCATCGGCACTGTCCTTGACCAAGCAGAGACTGCGGGGGCGAGACTGGTTGT
    GTAGAACCCGTAGCCGTGACAGGAACTGGTTCGTCTCTGACGCCCCCGCTCTGACCAACA

    LeuAlaThrAlaThrProProGlySerValThrValProHisProAsnIleGluGluVal
241 GCTCGCCACCGCCACCCCTCCGGGCTCCGTCACTGTGCCCCATCCCAACATCGAGGAGGT
    CGAGCGGTGGCGGTGGGGAGGCCCGAGGCAGTGACACGGGGTAGGGTTGTAGCTCCTCCA

    AlaLeuSerThrThrGlyGluIleProPheTyrGlyLysAlaIleProLeuGluValIle
301 TGCTCTGTCCACCACCGGAGAGATCCCTTTTTACGGCAAGGCTATCCCCCTCGAAGTAAT
    ACGAGACAGGTGGTGGCCTCTCTAGGGAAAAATGCCGTTCCGATAGGGGGAGCTTCATTA

    LysGlyGlyArgHisLeuIlePheCysHisSerLysLysLysCysAspGluLeuAlaAla
361 CAAGGGGGGGAGACATCTCATCTTCTGTCATTCAAAGAAGAAGTGCGACGAACTCGCCGC
    GTTCCCCCCCTCTGTAGAGTAGAAGACAGTAAGTTTCTTCTTCACGCTGCTTGAGCGGCG

    LysLeuValAlaLeuGlyIleAsnAlaValAlaTyrTyrArgGlyLeuAspValSerVal
421 AAAGCTGGTCGCATTGGGCATCAATGCCGTGGCCTACTACCGCGGTCTTGACGTGTCCGT
    TTTCGACCAGCGTAACCCGTAGTTACGGCACCGGATGATGGCGCCAGAACTGCACAGGCA

    IleProThrSerGlyAspValValValValAlaThrAspAlaLeuMetThrGlyTyrThr
481 CATCCCGACCAGCGGCGATGTTGTCGTCGTGGCAACCGATGCCCTCATGACCGGCTATAC
    GTAGGGCTGGTCGCCGCTACAACAGCAGCACCGTTGGCTACGGGAGTACTGGCCGATATG

    GlyAspPheAspSerValIleAspTyrAsnThrCysValThrGlnThrValAspPheSer
541 CGGCGACTTCGACTCGGTGATAGACTACAATACGTGTGTCACCCAGACAGTCGATTTCAG
    GCCGCTGAAGCTGAGCCACTATCTGATGTTATGCACACAGTGGGTCTGTCAGCTAAAGTC

    LeuAspProThrPheThrIleGluThrIleThrLeuProGlnAspAlaValSerArgThr
601 CCTTGACCCTACCTTCACCATTGAGACAATCACGCTCCCCCAGGATGCTGTCTCCCGCAC
    GGAACTGGGATGGAAGTGGTAACTCTGTTAGTGCGAGGGGGTCCTACGACAGAGGGCGTG

    GlnArgArgGlyArgThrGlyArgGlyLysProGlyIleTyrArgPheValAlaProGly
661 TCAACGTCGGGGCAGGACTGGCAGGGGGAAGCCAGGCATCTACAGATTTGTGGCACCGGG
    AGTTGCAGCCCCGTCCTGACCGTCCCCCTTCGGTCCGTAGATGTCTAAACACCGTGGCCC

    GluArgProSerGlyMetPheAspSerSerValLeuCysGluCysTyrAspAlaGlyCys
721 GGAGCGCCCCTCCGGCATGTTCGACTCGTCCGTCCTCTGTGAGTGCTATGACGCAGGCTG
    CCTCGCGGGGAGGCCGTACAAGCTGAGCAGGCAGGAGACACTCACGATACTGCGTCCGAC

    AlaTrpTyrGluLeuThrProAlaGluThrThrValArgLeuArgAlaTyrMetAsnThr
781 TGCTTGGTATGAGCTCACGCCCGCCGAGACTACAGTTAGGCTACGAGCGTACATGAACAC
    ACGAACCATACTCGAGTGCGGGCGGCTCTGATGTCAATCCGATGCTCGCATGTACTTGTG

    ProGlyLeuProValCysGlnAspHisLeuGluPheTrpGluGlyValPheThrGlyLeu
841 CCCGGGGCTTCCCGTGTGCCAGGACCATCTTGAATTTTGGGAGGGCGTCTTTACAGGCCT
    GGGCCCCGAAGGGCACACGGTCCTGGTAGAACTTAAAACCCTCCCGCAGAAATGTCCGGA

    ThrHisIleAspAlaHisPheLeuSerGlnThrLysGlnSerGlyGluAsnLeuProTyr
901 CACTCATATAGATGCCCACTTTCTATCCCAGACAAAGCAGAGTGGGGAGAACCTTCCTTA
    GTGAGTATATCTACGGGTGAAAGATAGGGTCTGTTTCGTCTCACCCCTCTTGGAAGGAAT
```

```
     LeuValAlaTyrGlnAlaThrValCysAlaArgAlaGlnAlaProProProSerTrpAsp
961  CCTGGTAGCGTACCAAGCCACCGTGTGCGCTAGGGCTCAAGCCCCTCCCCCATCGTGGGA
     GGACCATCGCATGGTTCGGTGGCACACGCGATCCCGAGTTCGGGGAGGGGGTAGCACCCT

     GlnMetTrpLysCysLeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeu
1021 CCAGATGTGGAAGTGTTTGATTCGCCTCAAGCCCACCCTCCATGGGCCAACACCCCTGCT
     GGTCTACACCTTCACAAACTAAGCGGAGTTCGGGTGGGAGGTACCCGGTTGTGGGGACGA

     TyrArgLeuGlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIle
1081 ATACAGACTGGGCGCTGTTCAGAATGAAATCACCCTGACGCACCCAGTCACCAAATACAT
     TATGTCTGACCCGCGACAAGTCTTACTTTAGTGGGACTGCGTGGGTCAGTGGTTTATGTA

     MetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyGly
1141 CATGACATGCATGTCGGCCGACCTGGAGGTCGTCACGAGCACCTGGGTGCTCGTTGGCGG
     GTACTGTACGTACAGCCGGCTGGACCTCCAGCAGTGCTCGTGGACCCACGAGCAACCGCC

     ValLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValValIleValGlyArg
1201 CGTCCTGGCTGCTTTGGCCGCGTATTGCCTGTCAACAGGCTGCGTGGTCATAGTGGGCAG
     GCAGGACCGACGAAACCGGCGCATAACGGACAGTTGTCCGACGCACCAGTATCACCCGTC

     ValValLeuSerGlyLysProAlaIleIleProAspArgGluValLeuTyrArgGluPhe
1261 GGTCGTCTTGTCCGGGAAGCCGGCAATCATACCTGACAGGGAAGTCCTCTACCGAGAGTT
     CCAGCAGAACAGGCCCTTCGGCCGTTAGTATGGACTGTCCCTTCAGGAGATGGCTCTCAA

     AspGluMetGluGluCysSerGlnHisLeuProTyrIleGluGlnGlyMetMetLeuAla
1321 CGATGAGATGGAAGAGTGCTCTCAGCACTTACCGTACATCGAGCAAGGGATGATGCTCGC
     GCTACTCTACCTTCTCACGAGAGTCGTGAATGGCATGTAGCTCGTTCCCTACTACGAGCG

     GluGlnPheLysGlnLysAlaLeuGlyLeuLeuGlnThrAlaSerArgGlnAlaGluVal
1381 CGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCCTGCAGACCGCGTCCCGTCAGGCAGAGGT
     GCTCGTCAAGTTCGTCTTCCGGGAGCCGGAGGACGTCTGGCGCAGGGCAGTCCGTCTCCA

     IleAlaProAlaValGlnThrAsnTrpGlnLysLeuGluThrPheTrpAlaLysHisMet
1441 TATCGCCCCTGCTGTCCAGACCAACTGGCAAAAACTCGAGACCTTCTGGGCGAAGCATAT
     ATAGCGGGGACGACAGGTCTGGTTGACCGTTTTTGAGCTCTGGAAGACCCGCTTCGTATA

     TrpAsnPheIleSerGlyIleGlnTyrLeuAlaGlyLeuSerThrLeuProGlyAsnPro
1501 GTGGAACTTCATCAGTGGGATACAATACTTGGCGGGCTTGTCAACGCTGCCTGGTAACCC
     CACCTTGAAGTAGTCACCCTATGTTATGAACCGCCCGAACAGTTGCGACGGACCATTGGG

     AlaIleAlaSerLeuMetAlaPheThrAlaAlaValThrSerProLeuThrThrSerGln
1561 CGCCATTGCTTCATTGATGGCTTTTACAGCTGCTGTCACCAGCCCACTAACCACTAGCCA
     GCGGTAACGAAGTAACTACCGAAAATGTCGACGACAGTGGTCGGGTGATTGGTGATCGGT

     ThrLeuLeuPheAsnIleLeuGlyGlyTrpValAlaAlaGlnLeuAlaAlaProGlyAla
1621 AACCCTCCTCTTCAACATATTGGGGGGGTGGGTGGCTGCCCAGCTCGCCGCCCCCGGTGC
     TTGGGAGGAGAAGTTGTATAACCCCCCCACCCACCGACGGGTCGAGCGGCGGGGGCCACG

     AlaThrAlaPheValGlyAlaGlyLeuAlaGlyAlaAlaIleGlySerValGlyLeuGly
1681 CGCTACTGCCTTTGTGGGCGCTGGCTTAGCTGGCGCCGCCATCGGCAGTGTTGGACTGGG
     GCGATGACGGAAACACCCGCGACCGAATCGACCGCGGCGGTAGCCGTCACAACCTGACCC

     LysValLeuIleAspIleLeuAlaGlyTyrGlyAlaGlyValAlaGlyAlaLeuValAla
1741 GAAGGTCCTCATAGACATCCTTGCAGGGTATGGCGCGGGCGTGGCGGGAGCTCTTGTGGC
     CTTCCAGGAGTATCTGTAGGAACGTCCCATACCGCGCCCGCACCGCCCTCGAGAACACCG

     PheLysIleMetSerGlyGluValProSerThrGluAspLeuValAsnLeuLeuProAla
1801 ATTCAAGATCATGAGCGGTGAGGTCCCCTCCACGGAGGACCTGGTCAATCTACTGCCCGC
     TAAGTTCTAGTACTCGCCACTCCAGGGGAGGTGCCTCCTGGACCAGTTAGATGACGGGCG

     IleLeuSerProGlyAlaLeuValValGlyValValCysAlaAlaIleLeuArgArgHis
1861 CATCCTCTCGCCCGGAGCCCTCGTAGTCGGCGTGGTCTGTGCAGCAATACTGCGCCGGCA
     GTAGGAGAGCGGGCCTCGGGAGCATCAGCCGCACCAGACACGTCGTTATGACGCGGCCGT

     ValGlyProGlyGluGlyAlaValGlnTrpMetAsnArgLeuIleAlaPheAlaSerArg
```

FIG. 14-2

```
1921  CGTTGGCCCGGGCGAGGGGGCAGTGCAGTGGATGAACCGGCTGATAGCCTTCGCCTCCCG
      GCAACCGGGCCCGCTCCCCCGTCACGTCACCTACTTGGCCGACTATCGGAAGCGGAGGGC

       GlyAsnHisValSerProThrHisTyrValProGluSerAspAlaAlaAlaArgValThr
1981  GGGGAACCATGTTTCCCCCACGCACTACGTGCCGGAGAGCGATGCAGCTGCCCGCGTCAC
      CCCCTTGGTACAAAGGGGGTGCGTGATGCACGGCCTCTCGCTACGTCGACGGGCGCAGTG

       AlaIleLeuSerSerLeuThrValThrGlnLeuLeuArgArgLeuHisGlnTrpIleSer
2041  TGCCATACTCAGCAGCCTCACTGTAACCCAGCTCCTGAGGCGACTGCACCAGTGGATAAG
      ACGGTATGAGTCGTCGGAGTGACATTGGGTCGAGGACTCCGCTGACGTGGTCACCTATTC

       SerGluCysThrThrProCysSerGlySerTrpLeuArgAspIleTrpAspTrpIleCys
2101  CTCGGAGTGTACCACTCCATGCTCCGGTTCCTGGCTAAGGGACATCTGGGACTGGATATG
      GAGCCTCACATGGTGAGGTACGAGGCCAAGGACCGATTCCCTGTAGACCCTGACCTATAC

       GluValLeuSerAspPheLysThrTrpLeuLysAlaLysLeuMetProGlnLeuProGly
2161  CGAGGTGTTGAGCGACTTTAAGACCTGGCTAAAAGCTAAGCTCATGCCACAGCTGCCTGG
      GCTCCACAACTCGCTGAAATTCTGGACCGATTTTCGATTCGAGTACGGTGTCGACGGACC

       IleProPheValSerCysGlnArgGlyTyrLysGlyValTrpArgValAspGlyIleMet
2221  GATCCCCTTTGTGTCCTGCCAGCGCGGGTATAAGGGGGTCTGGCGAGTGGACGGCATCAT
      CTAGGGGAAACACAGGACGGTCGCGCCCATATTCCCCCAGACCGCTCACCTGCCGTAGTA

       HisThrArgCysHisCysGlyAlaGluIleThrGlyHisValLysAsnGlyThrMetArg
2281  GCACACTCGCTGCCACTGTGGAGCTGAGATCACTGGACATGTCAAAAACGGGACGATGAG
      CGTGTGAGCGACGGTGACACCTCGACTCTAGTGACCTGTACAGTTTTTGCCCTGCTACTC

       IleValGlyProArgThrCysArgAsnMetTrpSerGlyThrPheProIleAsnAlaTyr
2341  GATCGTCGGTCCTAGGACCTGCAGGAACATGTGGAGTGGGACCTTCCCCATTAATGCCTA
      CTAGCAGCCAGGATCCTGGACGTCCTTGTACACCTCACCCTGGAAGGGGTAATTACGGAT

       ThrThrGlyProCysThrProLeuProAlaProAsnTyrThrPheAlaLeuTrpArgVal
2401  CACCACGGGCCCCTGTACCCCCCTTCCTGCGCCGAACTACACGTTCGCGCTATGGAGGGT
      GTGGTGCCCGGGGACATGGGGGGAAGGACGCGGCTTGATGTGCAAGCGCGATACCTCCCA

       SerAlaGluGluTyrValGluIleArgGlnValGlyAspPheHisTyrValThrGlyMet
2461  GTCTGCAGAGGAATATGTGGAGATAAGGCAGGTGGGGGACTTCCACTACGTGACGGGTAT
      CAGACGTCTCCTTATACACCTCTATTCCGTCCACCCCCTGAAGGTGATGCACTGCCCATA

       ThrThrAspAsnLeuLysCysProCysGlnValProSerProGluPhePheThrGlu
2521  GACTACTGACAATCTCAAATGCCCGTGCCAGGTCCCATCGCCCGAATTTTTCACAGAAT
      CTGATGACTGTTAGAGTTTACGGCACGGTCCAGGGTAGCGGGCTTAAAAAGTGTCTTA
```

FIG. 14-3

FIG. 15 Translation of DNA 33c

```
     AlaValAspPheIleProValGluAsnLeuGluThrThrMetArgSerProValPheThr
1    GGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATGAGGTCCCCGGTGTTCAC
     CCGCCACCTGAAATAGGGACACCTCTTGGATCTCTGTTGGTACTCCAGGGGCCACAAGTG

     AspAsnSerSerProProValValProGlnSerPheGlnValAlaHisLeuHisAlaPro
61   GGATAACTCCTCTCCACCAGTAGTGCCCCAGAGCTTCCAGGTGGCTCACCTCCATGCTCC
     CCTATTGAGGAGAGGTGGTCATCACGGGGTCTCGAAGGTCCACCGAGTGGAGGTACGAGG

     ThrGlySerGlyLysSerThrLysValProAlaAlaTyrAlaAlaGlnGlyTyrLysVal
121  CACAGGCAGCGGCAAAAGCACCAAGGTCCCGGCTGCATATGCAGCTCAGGGCTATAAGGT
     GTGTCCGTCGCCGTTTTCGTGGTTCCAGGGCCGACGTATACGTCGAGTCCCGATATTCCA

     LeuValLeuAsnProSerValAlaAlaThrLeuGlyPheGlyAlaTyrMetSerLysAla
181  GCTAGTACTCAACCCCTCTGTTGCTGCAACACTGGGCTTTGGTGCTTACATGTCCAAGGC
     CGATCATGAGTTGGGGAGACAACGACGTTGTGACCCGAAACCACGAATGTACAGGTTCCG

     ----------Overlap with 40b------------------------------------
     HisGlyIleAspProAsnIleArgThrGlyValArgThrIleThrThrGlySerProIle
241  TCATGGGATCGATCCTAACATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCCCCAT
     AGTACCCTAGCTAGGATTGTAGTCCTGGCCCCACTCTTGTTAATGGTGACCGTCGGGGTA

     ThrTyrSerThrTyrGlyLysPheLeuAlaAspGlyGlyCysSerGlyGlyAlaTyrAsp
301  CACGTACTCCACCTACGGCAAGTTCCTTGCCGACGGCGGGTGCTCGGGGGGCGCTTATGA
     GTGCATGAGGTGGATGCCGTTCAAGGAACGGCTGCCGCCCACGAGCCCCCCGCGAATACT

     IleIleIleCysAspGluCysHisSerThrAspAlaThrSerIleLeuGlyIleGlyThr
361  CATAATAATTTGTGACGAGTGCCACTCCACGGATGCCACATCCATCTTGGGCATTGGCAC
     GTATTATTAAACACTGCTCACGGTGAGGTGCCTACGGTGTAGGTAGAACCCGTAACCGTG

     ValLeuAspGlnAlaGluThrAlaGlyAlaArgLeuValValLeuAlaThrAlaThrPro
421  TGTCCTTGACCAAGCAGAGACTGCGGGGGCGAGACTGGTTGTGCTCGCCACCGCCACCCC
     ACAGGAACTGGTTCGTCTCTGACGCCCCCGCTCTGACCAACACGAGCGGTGGCGGTGGGG

     ProGlySerValThrValProHisProAsnIleGluGluValAlaLeuSerThrThrGly
481  TCCGGGCTCCGTCACTGTGCCCCATCCCAACATCGAGGAGGTTGCTCTGTCCACCACCGG
     AGGCCCGAGGCAGTGACACGGGGTAGGGTTGTAGCTCCTCCAACGAGACAGGTGGTGGCC

     GluIleProPheTyrGlyLysAlaIleProLeuGluValIleLysGlyGlyArgHisLeu
541  AGAGATCCCTTTTTACGGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGGAGACATCT
     TCTCTAGGGAAAAATGCCGTTCCGATAGGGGGAGCTTCATTAGTTCCCCCCCTCTGTAGA

     IlePheCysHisSerLysLysLysCysAspGluLeuAlaAlaLysLeuValAlaLeuGly
601  CATCTTCTGTCATTCAAAGAAGAAGTGCGACGAACTCGCCGCAAAGCTGGTCGCATTGGG
     GTAGAAGACAGTAAGTTTCTTCTTCACGCTGCTTGAGCGGCGTTTCGACCAGCGTAACCC

     IleAsnAlaValAlaTyrTyrArgGlyLeuAspValSerValIleProThrSerGlyAsp
661  CATCAATGCCGTGGCCTACTACCGCGGTCTTGACGTGTCCGTCATCCCGACCAGCGGCGA
     GTAGTTACGGCACCGGATGATGGCGCCAGAACTGCACAGGCAGTAGGGCTGGTCGCCGCT

     ValValValValAlaThrAspAlaLeuMetThrGlyTyrThrGlyAspPheAspSerVal
721  TGTTGTCGTCGTGGCAACCGATGCCCTCATGACCGGCTATACCGGCGACTTCGACTCGGT
     ACAACAGCAGCACCGTTGGCTACGGGAGTACTGGCCGATATGGCCGCTGAAGCTGAGCCA

     IleAspCysAsnThrCys
781  GATAGACTGCAATACGTGTG
     CTATCTGACGTTATGCACAC
```

FIG. 16 Translation of DNA 8h

```
    ProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIlePro
  1 CTCCCTGCACTTGCGGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTC
    GAGGGACGTGAACGCCGAGGAGCCTGGAAATGGACCAGTGCTCCGTGCGGCTACAGTAAG

    ValArgArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeu
 61 CCGTGCGCCGGCGGGGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACT
    GGCACGCGGCCGCCCCACTATCGTCCCCGTCGGACGACAGCGGGGCCGGGTAAAGGATGA

    LysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArg
121 TGAAAGGCTCCTCGGGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTA
    ACTTTCCGAGGAGCCCCCCAGGCGACAACACGGGGCGCCCCGTGCGGCACCCGTATAAAT

                                               ------------Overlap with
    AlaAlaValCysThrArgGlyValAlaLysAlaValAspPheIleProValGluAsnLeu
181 GGGCCGCGGTGTGCACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACC
    CCCGGCGCCACACGTGGGCACCTCACCGATTCCGCCACCTGAAATAGGGACACCTCTTGG

    33c--------------------------------------
    GluThrThrMetArgSerProValPheThrAspAsnSer
241 TAGAGACAACCATGAGGTCCCCGGTGTTCACGGATAACTCCTC
    ATCTCTGTTGGTACTCCAGGGGCCACAAGTGCCTATTGAGGAG
```

FIG. 17 Translation of DNA 7e

```
    GlyTrpArgLeuLeuAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGly
  1 GGGGTGGAGGTTGCTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGG
    CCCCACCTCCAACGACCGCGGGTAGTGCCGCATGCGGGTCGTCTGTTCCCCGGAGGATCC

    CysIleIleThrSerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIle
 61 GTGCATAATCACCAGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGAT
    CACGTATTAGTGGTCGGATTGACCGGCCCTGTTTTTGGTTCACCTCCCACTCCAGGTCTA

    ValSerThrAlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrVal
121 TGTGTCAACTGCTGCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGT
    ACACAGTTGACGACGGGTTTGGAAGGACCGTTGCACGTAGTTACCCCACACGACCTGACA

    TyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyr
181 CTACCACGGGGCCGGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTA
    GATGGTGCCCCGGCCTTGCTCCTGGTAGCGCAGTGGGTTCCCAGGACAGTAGGTCTACAT

    ThrAsnValAspGlnAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThr
241 TACCAATGTAGACCAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGAC
    ATGGTTACATCTGGTTCTGGAACACCCGACCGGGCGAGGCGTTCCATCGGCGAGTAACTG

    ----------Overlap with 8h---------------------
    ProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHis
301 ACCCTGCACTTGCGGCTCCTCGGACCTTTACCTGGTCACGAGGCACG
    TGGGACGTGAACGCCGAGGAGCCTGGAAATGGACCAGTGCTCCGTGC
```

FIG. 18 Translation of DNA 14c

```
     ------------------------------------------------------------
      AsnMetTrpSerGlyThrPheProIleAsnAlaTyrThrThrGlyProCysThrProLeu
  1  GAACATGTGGAGTGGACCTTCCCCATTAATGCCTACACCACGGGCCCCTGTACCCCCCT
     CTTGTACACCTCACCCTGGAAGGGGTAATTACGGATGTGGTGCCCGGGGACATGGGGGA

     -------------Overlap with 25c-------------------------------
      ProAlaProAsnTyrThrPheAlaLeuTrpArgValSerAlaGluGluTyrValGluIle
 61  TCCTGCGCCGAACTACACGTTCGCGCTATGGAGGGTGTCTGCAGAGGAATACGTGGAGAT
     AGGACGCGGCTTGATGTGCAAGCGCGATACCTCCCACAGACGTCTCCTTATGCACCTCTA

     ------------------------------------------------------------
      ArgGlnValGlyAspPheHisTyrValThrGlyMetThrThrAspAsnLeuLysCysPro
121  AAGGCAGGTGGGGGACTTCCACTACGTGACGGGTATGACTACTGACAATCTTAAATGCCC
     TTCCGTCCACCCCCTGAAGGTGATGCACTGCCCATACTGATGACTGTTAGAATTTACGGG

     ----------------------------------
      CysGlnValProSerProGluPhePheThrGluLeuAspGlyValArgLeuHisArgPhe
181  GTGCCAGGTCCCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCCTACATAGGTT
     CACGGTCCAGGGTAGCGGGCTTAAAAAGTGTCTTAACCTGCCCCACGCGGATGTATCCAA

      AlaProProCysLysProLeuLeuArgGluGluValSerPheArgValGlyLeuHisGlu
241  TGCGCCCCCCTGCAAGCCCTTGCTGCGGGAGGAGGTATCATTCAGAGTAGGACTCCACGA
     ACGCGGGGGGACGTTCGGGAACGACGCCCTCCTCCATAGTAAGTCTCATCCTGAGGTGCT

      TyrProValGlySerGlnLeuProCysGluProGluProAspValAlaValLeuThrSer
301  ATACCCGGTAGGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTGGCCGTGTTGACGTC
     TATGGGCCATCCCAGCGTTAATGGAACGCTCGGGCTTGGCCTGCACCGGCACAACTGCAG

      MetLeuThrAspProSerHisIleThrAlaGluAlaAlaGlyArgArgLeuAlaArgGly
361  CATGCTCACTGATCCCTCCCATATAACAGCAGAGGCGGCCGGGCGAAGGTTGGCGAGGGG
     GTACGAGTGACTAGGGAGGGTATATTGTCGTCTCCGCCGGCCCGCTTCCAACCGCTCCCC

      SerProProSerValAlaSerSerSerAlaSerGlnLeuSerAlaProSerLeuLysAla
421  ATCACCCCCCTCTGTGGCCAGCTCCTCGGCTAGCCAGCTATCCGCTCCATCTCTCAAGGC
     TAGTGGGGGGAGACACCGGTCGAGGAGCCGATCGGTCGATAGGCGAGGTAGAGAGTTCCG

      ThrCysThrAlaAsnHisAspSerProAsp
481  AACTTGCACCGCTAACCATGACTCCCCTGAT
     TTGAACGTGGCGATTGGTACTGAGGGGACTA
```

FIG. 19 Translation of DNA 8f

```
                                   Overlap with 14c
     SerSerSerAlaSerGlnLeuSerAlaProSerLeuLysAlaThrCysThrAlaAsnHis
  1  AGCTCCTCGGCTAGCCAGCTATCCGCTCCATCTCTCAAGGCAACTTGCACCGCTAACCAT
     TCGAGGAGCCGATCGGTCGATAGGCGAGGTAGAGAGTTCCGTTGAACGTGGCGATTGGTA

     AspSerProAspAlaGluLeuIleGluAlaAsnLeuLeuTrpArgGlnGluMetGlyGly
 61  GACTCCCCTGATGCTGAGCTCATAGAGGCCAACCTCCTATGGAGGCAGGAGATGGGCGGC
     CTGAGGGGACTACGACTCGAGTATCTCCGGTTGGAGGATACCTCCGTCCTCTACCCGCCG

     AsnIleThrArgValGluSerGluAsnLysValValIleLeuAspSerPheAspProLeu
121  AACATCACCAGGGTTGAGTCAGAAAACAAAGTGGTGATTCTGGACTCCTTCGATCCGCTT
     TTGTAGTGGTCCCAACTCAGTCTTTTGTTTCACCACTAAGACCTGAGGAAGCTAGGCGAA

     ValAlaGluGluAspGluArgGluIleSerValProAlaGluIleLeuArgLysSerArg
181  GTGGCGGAGGAGGACGAGCGGGAGATCTCCGTACCCGCAGAAATCCTGCGGAAGTCTCGG
     CACCGCCTCCTCCTGCTCGCCCTCTAGAGGCATGGGCGTCTTTAGGACGCCTTCAGAGCC

     ArgPheAlaGlnAlaLeuProValTrpAlaArgProAspTyrAsnProProLeuValGlu
241  AGATTCGCCCAGGCCCTGCCCGTTTGGGCGCGGCCGGACTATAACCCCCCGCTAGTGGAG
     TCTAAGCGGGTCCGGGACGGGCAAACCCGCGCCGGCCTGATATTGGGGGGCGATCACCTC

     ThrTrpLysLysProAspTyrGluProProValValHisGlyCysProLeuProProPro
301  ACGTGGAAAAAGCCCGACTACGAACCACCTGTGGTCCATGGCTGTCCGCTTCCACCTCCA
     TGCACCTTTTTCGGGCTGATGCTTGGTGGACACCAGGTACCGACAGGCGAAGGTGGAGGT

     LysSerProProValPro
361  AAGTCCCCTCCTGTGCCG
     TTCAGGGGAGGACACGGC
```

FIG. 20 Translation of DNA 33f

```
     ValTrpAlaArgProAspTyrAsnProProLeuValGluThrTrpLysLysProAspTyr
  1  CGTTTGGGCGCGGCCGGACTATAACCCCCCGCTAGTGGAGACGTGGAAAAAACCCGACTA
     GCAAACCCGCGCCGGCCTGATATTGGGGGGCGATCACCTCTGCACCTTTTTTGGGCTGAT

                       Overlap with 8f
     GluProProValValHisGlyCysProLeuProProProLysSerProProValProPro
 61  CGAACCACCTGTGGTCCATGGCTGCCCGCTTCCACCTCCAAAGTCCCCTCCTGTGCCTCC
     GCTTGGTGGACACCAGGTACCGACGGGCGAAGGTGGAGGTTTCAGGGGAGGACACGGAGG

     ProArgLysLysArgThrValValLeuThrGluSerThrLeuSerThrAlaLeuAlaGlu
121  GCCTCGGAAGAAGCGGACGGTGGTCCTCACTGAATCAACCCTATCTACTGCCTTGGCCGA
     CGGAGCCTTCTTCGCCTGCCACCAGGAGTGACTTAGTTGGGATAGATGACGGAACCGGCT

     LeuAlaThrArgSerPheGlySerSerSerThrSerGlyIleThrGlyAspAsnThrThr
181  GCTCGCCACCAGAAGCTTTGGCAGCTCCTCAACTTCCGGCATTACGGGCGACAATACGAC
     CGAGCGGTGGTCTTCGAAACCGTCGAGGAGTTGAAGGCCGTAATGCCCGCTGTTATGCTG

     ThrSerSerGluProAlaProSerGlyCysProProAspSerAspAlaGluSerPhe
241  AACATCCTCTGAGCCCGCCCCTTCTGGCTGCCCCCCCGACTCCGACGCTGAGTCCTTTGC
     TTGTAGGAGACTCGGGCGGGGAAGACCGACGGGGGGGCTGAGGCTGCGACTCAGGAAACG
```

19/63

FIG. 21 Translation of DNA 33g

```
    ------------------------------------------------------------
    AlaSerArgSerPheGlySerSerSerThrSerGlyIleThrGlyAspAsnThrThrThr
  1 GCCTCCAGAAGCTTTGGCAGCTCCTCAACTTCCGGCATTACGGGCGACAATACGACAACA
    CGGAGGTCTTCGAAACCGTCGAGGAGTTGAAGGCCGTAATGCCCGCTGTTATGCTGTTGT

    -------------Overlap with 33f---------------------
    SerSerGluProAlaProSerGlyCysProProAspSerAspAlaGluSerTyrSerSer
 61 TCCTCTGAGCCCGCCCCTTCTGGCTGCCCCCCCGACTCCGACGCTGAGTCCTATTCCTCC
    AGGAGACTCGGGCGGGGAAGACCGACGGGGGGGCTGAGGCTGCGACTCAGGATAAGGAGG

    MetProProLeuGluGlyGluProGlyAspProAspLeuSerAspGlySerTrpSerThr
121 ATGCCCCCCCTGGAGGGGGAGCCTGGGGATCCGGATCTTAGCGACGGGTCATGGTCAACG
    TACGGGGGGGACCTCCCCCTCGGACCCCTAGGCCTAGAATCGCTGCCCAGTACCAGTTGC

    ValSerSerGluAlaAsnAlaGluAspValValCysCysSerMetSerTyrSerTrpThr
181 GTCAGTAGTGAGGCCAACGCGGAGGATGTCGTGTGCTGCTCAATGTCTTACTCTTGGACA
    CAGTCATCACTCCGGTTGCGCCTCCTACAGCACACGACGAGTTACAGAATGAGAACCTGT

    GlyAlaLeuValThrProCysAlaAlaGluGluGlnLysLeuProIleAsnAlaLeuSer
241 GGCGCACTCGTCACCCCGTGCGCCGCGGAAGAACAGAAACTGCCCATCAATGCACTAAGC
    CCGCGTGAGCAGTGGGGCACGCGGCGCCTTCTTGTCTTTGACGGGTAGTTACGTGATTCG

    AsnSerLeuLeuArgHisHisAsnLeuValTyrSerThrThrSerArgSer
301 AACTCGTTGCTACGTCACCACAATTTGGTGTATTCCACCACCTCACGCAGTG
    TTGAGCAACGATGCAGTGGTGTTAAACCACATAAGGTGGTGGAGTGCGTCAC
```

FIG. 22 Translation of DNA 7f

```
    GlyThrTyrValTyrAsnHisLeuThrProLeuArgAspTrpAlaHisAsnGlyLeuArg
  1 GGCACCTATGTTTATAACCATCTCACTCCTCTTCGGGACTGGGCGCACAACGGCTTGCGA
    CCGTGGATACAAATATTGGTAGAGTGAGGAGAAGCCCTGACCCGCGTGTTGCCGAACGCT

    AspLeuAlaValAlaValGluProValValPheSerGlnMetGluThrLysLeuIleThr
 61 GATCTGGCCGTGGCTGTAGAGCCAGTCGTCTTCTCCCAAATGGAGACCAAGCTCATCACG
    CTAGACCGGCACCGACATCTCGGTCAGCAGAAGAGGGTTTACCTCTGGTTCGAGTAGTGC

    TrpGlyAlaAspThrAlaAlaCysGlyAspIleIleAsnGlyLeuProValSerAlaArg
121 TGGGGGGCAGATACCGCCGCGTGCGGTGACATCATCAACGGCTTGCCTGTTTCCGCCCGC
    ACCCCCCGTCTATGGCGGCGCACGCCACTGTAGTAGTTGCCGAACGGACAAAGGCGGGCG

                                                  --------------
    ArgGlyArgGluIleLeuLeuGlyProAlaAspGlyMetValSerLysGlyTrpArgLeu
181 AGGGGCCGGGAGATACTGCTCGGGCCAGCCGATGGAATGGTCTCCAAGGGTTGGAGGTTG
    TCCCCGGCCCTCTATGACGAGCCCGGTCGGCTACCTTACCAGAGGTTCCCAACCTCCAAC

    ------------------------------------------------------------
    LeuAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr
241 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC
    GACCGCGGGTAGTGCCGCATGCGGGTCGTCTGTTCCCCGGAGGATCCCACGTATTAGTGG

    ----------------Overlap with 7e-----------------------------
    SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla
301 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT
    TCGGATTGACCGGCCCTGTTTTTGGTTCACCTCCCACTCCAGGTCTAACACAGTTGACGA

    ------------------------------------------
    AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrp
361 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGG
    CGGGTTTGGAAGGACCGTTGCACGTAGTTACCCCACACGACC
```

FIG. 23 Translation of DNA 11b

```
     GlyGlyValValLeuValGlyLeuMetAlaLeuThrLeuSerProTyrTyrLysArgTyr
1    GGCGGTGTTGTTCTCGTCGGGTTGATGGCGCTGACTCTGTCACCATATTACAAGCGCTAT
     CCGCCACAACAAGAGCAGCCCAACTACCGCGACTGAGACAGTGGTATAATGTTCGCGATA

     IleSerTrpCysLeuTrpTrpLeuGlnTyrPheLeuThrArgValGluAlaGlnLeuHis
61   ATCAGCTGGTGCTTGTGGTGGCTTCAGTATTTTCTGACCAGAGTGGAAGCGCAACTGCAC
     TAGTCGACCACGAACACCACCGAAGTCATAAAAGACTGGTCTCACCTTCGCGTTGACGTG

     ValTrpIleProProLeuAsnValArgGlyGlyArgAspAlaValIleLeuLeuMetCys
121  GTGTGGATTCCCCCCCTCAACGTCCGAGGGGGGCGCGACGCCGTCATCTTACTCATGTGT
     CACACCTAAGGGGGGGAGTTGCAGGCTCCCCCCGCGCTGCGGCAGTAGAATGAGTACACA

     AlaValHisProThrLeuValPheAspIleThrLysLeuLeuLeuAlaValPheGlyPro
181  GCTGTACACCCGACTCTGGTATTTGACATCACCAAATTGCTGCTGGCCGTCTTCGGACCC
     CGACATGTGGGCTGAGACCATAAACTGTAGTGGTTTAACGACGACCGGCAGAAGCCTGGG

     LeuTrpIleLeuGlnAlaSerLeuLeuLysValProTyrPheValArgValGlnGlyLeu
241  CTTTGGATTCTTCAAGCCAGTTTGCTTAAAGTACCCTACTTTGTGCGCGTCCAAGGCCTT
     GAAACCTAAGAAGTTCGGTCAAACGAATTTCATGGGATGAAACACGCGCAGGTTCCGGAA

     LeuArgPheCysAlaLeuAlaArgLysMetIleGlyGlyHisTyrValGlnMetValIle
301  CTCCGGTTCTGCGCGTTAGCGCGGAAGATGATCGGAGGCCATTACGTGCAAATGGTCATC
     GAGGCCAAGACGCGCAATCGCGCCTTCTACTAGCCTCCGGTAATGCACGTTTACCAGTAG

                                 ------------------------------
     IleLysLeuGlyAlaLeuThrGlyThrTyrValTyrAsnHisLeuThrProLeuArgAsp
361  ATTAAGTTAGGGGCGCTTACTGGCACCTATGTTTATAACCATCTCACTCCTCTTCGGGAC
     TAATTCAATCCCCGCGAATGACCGTGGATACAAATATTGGTAGAGTGAGGAGAAGCCCTG

     ----------------Overlap with 7f ----------------------------
     TrpAlaHisAsnGlyLeuArgAspLeuAlaValAlaValGluProValValPheSerGln
421  TGGGCGCACAACGGCTTGCGAGATCTGGCCGTGGCTGTAGAGCCAGTCGTCTTCTCCCAA
     ACCCGCGTGTTGCCGAACGCTCTAGACCGGCACCGACATCTCGGTCAGCAGAAGAGGGTT

     ---------------------------
     MetGluThrLysLeuIleThrTrpGly
481  ATGGAGACCAAGCTCATCACGTGGGGGGC
     TACCTCTGGTTCGAGTAGTGCACCCCCCG
```

FIG. 24 Translation of DNA 14i

```
     GluTyrValValLeuLeuPheLeuLeuLeuAlaAspAlaArgValCysSerCysLeuTrp
  1  GGGAGTACGTCGTTCTCCTGTTCCTTCTGCTTGCAGACGCGCGTCTGCTCCTGCTTGT
     CCCTCATGCAGCAAGAGGACAAGGAAGACGAACGTCTGCGCGCAGACGAGGACGAACA

     MetMetLeuLeuIleSerGlnAlaGluAlaAlaLeuGluAsnLeuValIleLeuAsnAla
 61  GGATGATGCTACTCATATCCCAAGCGGAGGCGGCTTTGGAGAACCTCGTAATACTTAATG
     CCTACTACGATGAGTATAGGGTTCGCCTCCGCCGAAACCTCTTGGAGCATTATGAATTAC

     AlaSerLeuAlaGlyThrHisGlyLeuValSerPheLeuValPhePheCysPheAlaTrp
121  CAGCATCCCTGGCCGGGACGCACGGTCTTGTATCCTTCCTCGTGTTCTTCTGCTTTGCAT
     GTCGTAGGGACCGGCCCTGCGTGCCAGAACATAGGAAGGAGCACAAGAAGACGAAACGTA

     TyrLeuLysGlyLysTrpValProGlyAlaValTyrThrPheTyrGlyMetTrpProLeu
181  GGTATTTGAAGGGTAAGTGGGTGCCCGGAGCGGTCTACACCTTCTACGGGATGTGGCCTC
     CCATAAACTTCCCATTCACCCACGGGCCTCGCCAGATGTGGAAGATGCCCTACACCGGAG

     LeuLeuLeuLeuLeuAlaLeuProGlnArgAlaTyrAlaLeuAspThrGluValAlaAla
241  TCCTCCTGCTCCTGTTGGCGTTGCCCCAGCGGGCGTACGCGCTGGACACGGAGGTGGCCG
     AGGAGGACGAGGACAACCGCAACGGGGTCGCCCGCATGCGCGACCTGTGCCTCCACCGGC

                 --------------------Overlap with 11b ---------------
     SerCysGlyGlyValValLeuValGlyLeuMetAlaLeuThrLeuSerProTyrTyrLys
301  CGTCGTGTGGCGGTGTTGTTCTCGTCGGGTTGATGGCGCTGACTCTGTCACCATATTACA
     GCAGCACACCGCCACAACAAGAGCAGCCCAACTACCGCGACTGAGACAGTGGTATAATGT

     ---------------------------------------
     ArgTyrIleSerTrpCysLeuTrpTrpLeuGln
361  AGCGCTATATCAGCTGGTGCTTGTGGTGGCTTCAGAA
     TCGCGATATAGTCGACCACGAACACCACCGAAGTCTT
```

FIG. 25 Translation of DNA 39c

```
     ------------------------------------------------------------
     ProAlaProSerGlyCysProProAspSerAspAlaGluSerTyrSerSerMetProPro
  1  CCAGCCCCTTCTGGCTGCCCCCCCGACTCCGACGCTGAGTCCTATTCCTCCATGCCCCCC
     GGTCGGGGAAGACCGACGGGGGGGCTGAGGCTGCGACTCAGGATAAGGAGGTACGGGGGG

     ------------------------------------------------------------
     LeuGluGlyGluProGlyAspProAspLeuSerAspGlySerTrpSerThrValSerSer
 61  CTGGAGGGGGAGCCTGGGGATCCGGATCTTAGCGACGGGTCATGGTCAACAGTCAGTAGT
     GACCTCCCCCTCGGACCCCTAGGCCTAGAATCGCTGCCCAGTACCAGTTGTCAGTCATCA

     -------------------Overlap with 33g ------------------------
     GluAlaAsnAlaGluAspValValCysCysSerMetSerTyrSerTrpThrGlyAlaLeu
121  GAGGCCAACGCGGAGGATGTCGTGTGCTGCTCAATGTCCTACTCTTGGACAGGCGCACTC
     CTCCGGTTGCGCCTCCTACAGCACACGACGAGTTACAGGATGAGAACCTGTCCGCGTGAG

     ------------------------------------------------------------
     ValThrProCysAlaAlaGluGluGlnLysLeuProIleAsnAlaLeuSerAsnSerLeu
181  GTCACCCCGTGCGCCGCGGAAGAACAGAAACTGCCCATCAATGCACTGAGCAACTCGTTG
     CAGTGGGGCACGCGGCGCCTTCTTGTCTTTGACGGGTAGTTACGTGACTCGTTGAGCAAC

     ------------------------------------------
     LeuArgHisHisAsnLeuValTyrSerThrThrSerArgSerAlaCysGlnArgGlnLys
241  CTACGTCACCACAATTTGGTGTATTCCACCACCTCACGCAGTGCTTGCCAAAGGCAGAAG
     GATGCAGTGGTGTTAAACCACATAAGGTGGTGGAGTGCGTCACGAACGGTTTCCGTCTTC

     LysValThrPheAspArgLeuGlnValLeuAspSerHisTyrGlnAspValLeuLysGlu
301  AAAGTCACATTTGACAGACTGCAAGTTCTGGACAGCCATTACCAGGACGTACTCAAGGAG
     TTTCAGTGTAAACTGTCTGACGTTCAAGACCTGTCGGTAATGGTCCTGCATGAGTTCCTC

     ValLysAlaAlaAlaSerLysValLysAlaAsnPhe
361  GTTAAAGCAGCGGCGTCAAAAGTGAAGGCTAACTTC
     CAATTTCGTCGCCGCAGTTTTCACTTCCGATTGAAG
```

FIG. 26-1 COMBINED ORF OF DNAs 14i/11b/7f/7e/8h/33c/40b/37b/35/36/81/32/33b/25c/14c/8f/33f/33g/39c

```
     GluTyrValValLeuLeuPheLeuLeuLeuAlaAspAlaArgValCysSerCysLeuTrp
  1  GGGAGTACGTCGTTCTCCTGTTCCTTCTGCTTGCAGACGCGCGCGTCTGCTCCTGCTTGT
     CCCTCATGCAGCAAGAGGACAAGGAAGACGAACGTCTGCGCGCGCAGACGAGGACGAACA

     MetMetLeuLeuIleSerGlnAlaGluAlaAlaLeuGluAsnLeuValIleLeuAsnAla
 61  GGATGATGCTACTCATATCCCAAGCGGAGGCGGCTTTGGAGAACCTCGTAATACTTAATG
     CCTACTACGATGAGTATAGGGTTCGCCTCCGCCGAAACCTCTTGGAGCATTATGAATTAC

     AlaSerLeuAlaGlyThrHisGlyLeuValSerPheLeuValPhePheCysPheAlaTrp
121  CAGCATCCCTGGCCGGGACGCACGGTCTTGTATCCTTCCTCGTGTTCTTCTGCTTTGCAT
     GTCGTAGGGACCGGCCCTGCGTGCCAGAACATAGGAAGGAGCACAAGAAGACGAAACGTA

     TyrLeuLysGlyLysTrpValProGlyAlaValTyrThrPheTyrGlyMetTrpProLeu
181  GGTATTTGAAGGGTAAGTGGGTGCCCGGAGCGGTCTACACCTTCTACGGGATGTGGCCTC
     CCATAAACTTCCCATTCACCCACGGGCCTCGCCAGATGTGGAAGATGCCCTACACCGGAG

     LeuLeuLeuLeuLeuAlaLeuProGlnArgAlaTyrAlaLeuAspThrGluValAlaAla
241  TCCTCCTGCTCCTGTTGGCGTTGCCCCAGCGGGCGTACGCGCTGGACACGGAGGTGGCCG
     AGGAGGACGAGGACAACCGCAACGGGGTCGCCCGCATGCGCGACCTGTGCCTCCACCGGC

     SerCysGlyGlyValValLeuValGlyLeuMetAlaLeuThrLeuSerProTyrTyrLys
301  CGTCGTGTGGCGGTGTTGTTCTCGTCGGGTTGATGGCGCTGACTCTGTCACCATATTACA
     GCAGCACACCGCCACAACAAGAGCAGCCCAACTACCGCGACTGAGACAGTGGTATAATGT

     ArgTyrIleSerTrpCysLeuTrpTrpLeuGlnTyrPheLeuThrArgValGluAlaGln
361  AGCGCTATATCAGCTGGTGCTTGTGGTGGCTTCAGTATTTTCTGACCAGAGTGGAAGCGC
     TCGCGATATAGTCGACCACGAACACCACCGAAGTCATAAAAGACTGGTCTCACCTTCGCG

     LeuHisValTrpIleProProLeuAsnValArgGlyGlyArgAspAlaValIleLeuLeu
421  AACTGCACGTGTGGATTCCCCCCCTCAACGTCCGAGGGGGGCGCGACGCCGTCATCTTAC
     TTGACGTGCACACCTAAGGGGGGGAGTTGCAGGCTCCCCCCGCGCTGCGGCAGTAGAATG

     MetCysAlaValHisProThrLeuValPheAspIleThrLysLeuLeuLeuAlaValPhe
481  TCATGTGTGCTGTACACCCGACTCTGGTATTTGACATCACCAAATTGCTGCTGGCCGTCT
     AGTACACACGACATGTGGGCTGAGACCATAAACTGTAGTGGTTTAACGACGACCGGCAGA

     GlyProLeuTrpIleLeuGlnAlaSerLeuLeuLysValProTyrPheValArgValGln
541  TCGGACCCCTTTGGATTCTTCAAGCCAGTTTGCTTAAAGTACCCTACTTTGTGCGCGTCC
     AGCCTGGGGAAACCTAAGAAGTTCGGTCAAACGAATTTCATGGGATGAAACACGCGCAGG

     GlyLeuLeuArgPheCysAlaLeuAlaArgLysMetIleGlyGlyHisTyrValGlnMet
601  AAGGCCTTCTCCGGTTCTGCGCGTTAGCGCGGAAGATGATCGGAGGCCATTACGTGCAAA
     TTCCGGAAGAGGCCAAGACGCGCAATCGCGCCTTCTACTAGCCTCCGGTAATGCACGTTT

     ValIleIleLysLeuGlyAlaLeuThrGlyThrTyrValTyrAsnHisLeuThrProLeu
661  TGGTCATCATTAAGTTAGGGGCGCTTACTGGCACCTATGTTTATAACCATCTCACTCCTC
     ACCAGTAGTAATTCAATCCCCGCGAATGACCGTGGATACAAATATTGGTAGAGTGAGGAG

     ArgAspTrpAlaHisAsnGlyLeuArgAspLeuAlaValAlaValGluProValValPhe
721  TTCGGGACTGGGCGCACAACGGCTTGCGAGATCTGGCCGTGGCTGTAGAGCCAGTCGTCT
     AAGCCCTGACCCGCGTGTTGCCGAACGCTCTAGACCGGCACCGACATCTCGGTCAGCAGA

     SerGlnMetGluThrLysLeuIleThrTrpGlyAlaAspThrAlaAlaCysGlyAspIle
781  TCTCCCAAATGGAGACCAAGCTCATCACGTGGGGGGCAGATACCGCCGCGTGCGGTGACA
     AGAGGGTTTACCTCTGGTTCGAGTAGTGCACCCCCCGTCTATGGCGGCGCACGCCACTGT

     IleAsnGlyLeuProValSerAlaArgArgGlyArgGluIleLeuLeuGlyProAlaAsp
841  TCATCAACGGCTTGCCTGTTTCCGCCCGCAGGGGCCGGGAGATACTGCTCGGGCCAGCCG
     AGTAGTTGCCGAACGGACAAAGGCGGGCGTCCCCGGCCCTCTATGACGAGCCCGGTCGGC

     GlyMetValSerLysGlyTrpArgLeuLeuAlaProIleThrAlaTyrAlaGlnGlnThr
```

```
901   ATGGAATGGTCTCCAAGGGGTGGAGGTTGCTGGCGCCCATCACGGCGTACGCCCAGCAGA
      TACCTTACCAGAGGTTCCCCACCTCCAACGACCGCGGGTAGTGCCGCATGCGGGTCGTCT

       ArgGlyLeuLeuGlyCysIleIleThrSerLeuThrGlyArgAspLysAsnGlnValGlu
961   CAAGGGGCCTCCTAGGGTGCATAATCACCAGCCTAACTGGCCGGGACAAAAACCAAGTGG
      GTTCCCCGGAGGATCCCACGTATTAGTGGTCGGATTGACCGGCCCTGTTTTTGGTTCACC

       GlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThrCysIleAsnGly
1021  AGGGTGAGGTCCAGATTGTGTCAACTGCTGCCCAAACCTTCCTGGCAACGTGCATCAATG
      TCCCACTCCAGGTCTAACACAGTTGACGACGGGTTTGGAAGGACCGTTGCACGTAGTTAC

       ValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyPro
1081  GGGTGTGCTGGACTGTCTACCACGGGGCCGGAACGAGGACCATCGCGTCACCCAAGGGTC
      CCCACACGACCTGACAGATGGTGCCCCGGCCTTGCTCCTGGTAGCGCAGTGGGTTCCCAG

       ValIleGlnMetTyrThrAsnValAspGlnAspLeuValGlyTrpProAlaProGlnGly
1141  CTGTCATCCAGATGTATACCAATGTAGACCAAGACCTTGTGGGCTGGCCCGCTCCGCAAG
      GACAGTAGGTCTACATATGGTTACATCTGGTTCTGGAACACCCGACCGGGCGAGGCGTTC

       SerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHis
1201  GTAGCCGCTCATTGACACCCTGCACTTGCGGCTCCTCGGACCTTTACCTGGTCACGAGGC
      CATCGGCGAGTAACTGTGGGACGTGAACGCCGAGGAGCCTGGAAATGGACCAGTGCTCCG

       AlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeuLeuSerProArg
1261  ACGCCGATGTCATTCCCGTGCGCCGGCGGGGTGATAGCAGGGGCAGCCTGCTGTCGCCCC
      TGCGGCTACAGTAAGGGCACGCGGCCGCCCCACTATCGTCCCCGTCGGACGACAGCGGGG

       ProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAla
1321  GGCCCATTTCCTACTTGAAAGGCTCCTCGGGGGGTCCGCTGTTGTGCCCCGCGGGGCACG
      CCGGGTAAAGGATGAACTTTCCGAGGAGCCCCCCAGGCGACAACACGGGGCGCCCCGTGC

       ValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAlaValAspPheIle
1381  CCGTGGGCATATTTAGGGCCGCGGTGTGCACCCGTGGAGTGGCTAAGGCGGTGGACTTTA
      GGCACCCGTATAAATCCCGGCGCCACACGTGGGCACCTCACCGATTCCGCCACCTGAAAT

       ProValGluAsnLeuGluThrThrMetArgSerProValPheThrAspAsnSerSerPro
1441  TCCCTGTGGAGAACCTAGAGACAACCATGAGGTCCCCGGTGTTCACGGATAACTCCTCTC
      AGGGACACCTCTTGGATCTCTGTTGGTACTCCAGGGGCCACAAGTGCCTATTGAGGAGAG

       ProValValProGlnSerPheGlnValAlaHisLeuHisAlaProThrGlySerGlyLys
1501  CACCAGTAGTGCCCCAGAGCTTCCAGGTGGCTCACCTCCATGCTCCCACAGGCAGCGGCA
      GTGGTCATCACGGGGTCTCGAAGGTCCACCGAGTGGAGGTACGAGGGTGTCCGTCGCCGT

       SerThrLysValProAlaAlaTyrAlaAlaGlnGlyTyrLysValLeuValLeuAsnPro
1561  AAAGCACCAAGGTCCCGGCTGCATATGCAGCTCAGGGCTATAAGGTGCTAGTACTCAACC
      TTTCGTGGTTCCAGGGCCGACGTATACGTCGAGTCCCGATATTCCACGATCATGAGTTGG

       SerValAlaAlaThrLeuGlyPheGlyAlaTyrMetSerLysAlaHisGlyIleAspPro
1621  CCTCTGTTGCTGCAACACTGGGCTTTGGTGCTTACATGTCCAAGGCTCATGGGATCGATC
      GGAGACAACGACGTTGTGACCCGAAACCACGAATGTACAGGTTCCGAGTACCCTAGCTAG

       AsnIleArgThrGlyValArgThrIleThrThrGlySerProIleThrTyrSerThrTyr
1681  CTAACATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTCCACCT
      GATTGTAGTCCTGGCCCCACTCTTGTTAATGGTGACCGTCGGGGTAGTGCATGAGGTGGA

       GlyLysPheLeuAlaAspGlyGlyCysSerGlyGlyAlaTyrAspIleIleIleCysAsp
1741  ACGGCAAGTTCCTTGCCGACGGCGGGTGCTCGGGGGGCGCTTATGACATAATAATTTGTG
      TGCCGTTCAAGGAACGGCTGCCGCCCACGAGCCCCCCGCGAATACTGTATTATTAAACAC

       GluCysHisSerThrAspAlaThrSerIleLeuGlyIleGlyThrValLeuAspGlnAla
1801  ACGAGTGCCACTCCACGGATGCCACATCCATCTTGGGCATCGGCACTGTCCTTGACCAAG
      TGCTCACGGTGAGGTGCCTACGGTGTAGGTAGAACCCGTAGCCGTGACAGGAACTGGTTC

       GluThrAlaGlyAlaArgLeuValValLeuAlaThrAlaThrProProGlySerValThr
1861  CAGAGACTGCGGGGGCGAGACTGGTTGTGCTCGCCACCGCCACCCCTCCGGGCTCCGTCA
      GTCTCTGACGCCCCCGCTCTGACCAACACGAGCGGTGGCGGTGGGGAGGCCCGAGGCAGT
```

FIG. 26-2

ValProHisProAsnIleGluGluValAlaLeuSerThrThrGlyGluIleProPheTyr
1921 CTGTGCCCCATCCCAACATCGAGGAGGTTGCTCTGTCCACCACCGGAGAGATCCCTTTTT
GACACGGGGTAGGGTTGTAGCTCCTCCAACGAGACAGGTGGTGGCCTCTCTAGGGAAAAA

GlyLysAlaIleProLeuGluValIleLysGlyGlyArgHisLeuIlePheCysHisSer
1981 ACGGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGGAGACATCTCATCTTCTGTCATT
TGCCGTTCCGATAGGGGGAGCTTCATTAGTTCCCCCCCTCTGTAGAGTAGAAGACAGTAA

LysLysLysCysAspGluLeuAlaAlaLysLeuValAlaLeuGlyIleAsnAlaValAla
2041 CAAAGAAGAAGTGCGACGAACTCGCCGCAAAGCTGGTCGCATTGGGCATCAATGCCGTGG
GTTTCTTCTTCACGCTGCTTGAGCGGCGTTTCGACCAGCGTAACCCGTAGTTACGGCACC

TyrTyrArgGlyLeuAspValSerValIleProThrSerGlyAspValValValValAla
2101 CCTACTACCGCGGTCTTGACGTGTCCGTCATCCCGACCAGCGGCGATGTTGTCGTCGTGG
GGATGATGGCGCCAGAACTGCACAGGCAGTAGGGCTGGTCGCCGCTACAACAGCAGCACC

ThrAspAlaLeuMetThrGlyTyrThrGlyAspPheAspSerValIleAspCysAsnThr
2161 CAACCGATGCCCTCATGACCGGCTATACCGGCGACTTCGACTCGGTGATAGACTGCAATA
GTTGGCTACGGGAGTACTGGCCGATATGGCCGCTGAAGCTGAGCCACTATCTGACGTTAT

CysValThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThr
2221 CGTGTGTCACCCAGACAGTCGATTTCAGCCTTGACCCTACCTTCACCATTGAGACAATCA
GCACACAGTGGGTCTGTCAGCTAAAGTCGGAACTGGGATGGAAGTGGTAACTCTGTTAGT

LeuProGlnAspAlaValSerArgThrGlnArgArgGlyArgThrGlyArgGlyLysPro
2281 CGCTCCCCCAGGATGCTGTCTCCCGCACTCAACGTCGGGGCAGGACTGGCAGGGGGAAGC
GCGAGGGGGTCCTACGACAGAGGGCGTGAGTTGCAGCCCCGTCCTGACCGTCCCCCTTCG

GlyIleTyrArgPheValAlaProGlyGluArgProSerGlyMetPheAspSerSerVal
2341 CAGGCATCTACAGATTTGTGGCACCGGGGGAGCGCCCCTCCGGCATGTTCGACTCGTCCG
GTCCGTAGATGTCTAAACACCGTGGCCCCCTCGCGGGGAGGCCGTACAAGCTGAGCAGGC

LeuCysGluCysTyrAspAlaGlyCysAlaTrpTyrGluLeuThrProAlaGluThrThr
2401 TCCTCTGTGAGTGCTATGACGCAGGCTGTGCTTGGTATGAGCTCACGCCCGCCGAGACTA
AGGAGACACTCACGATACTGCGTCCGACACGAACCATACTCGAGTGCGGGCGGCTCTGAT

ValArgLeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGlu
2461 CAGTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCATCTTG
GTCAATCCGATGCTCGCATGTACTTGTGGGGCCCCGAAGGGCACACGGTCCTGGTAGAAC

PheTrpGluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThr
2521 AATTTTGGGAGGGCGTCTTTACAGGCCTCACTCATATAGATGCCCACTTTCTATCCCAGA
TTAAAACCCTCCCGCAGAAATGTCCGGAGTGAGTATATCTACGGGTGAAAGATAGGGTCT

LysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGlnAlaThrValCysAlaArg
2581 CAAAGCAGAGTGGGGAGAACCTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCTA
GTTTCGTCTCACCCCTCTTGGAAGGAATGGACCATCGCATGGTTCGGTGGCACACGCGAT

AlaGlnAlaProProProSerTrpAspGlnMetTrpLysCysLeuIleArgLeuLysPro
2641 GGGCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATTCGCCTCAAGC
CCCGAGTTCGGGGAGGGGGTAGCACCCTGGTCTACACCTTCACAAACTAAGCGGAGTTCG

ThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAlaValGlnAsnGluIleThr
2701 CCACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAAATCA
GGTGGGAGGTACCCGGTTGTGGGGACGATATGTCTGACCCGCGACAAGTCTTACTTTAGT

LeuThrHisProValThrLysTyrIleMetThrCysMetSerAlaAspLeuGluValVal
2761 CCCTGACGCACCCAGTCACCAAATACATCATGACATGCATGTCGGCCGACCTGGAGGTCG
GGGACTGCGTGGGTCAGTGGTTTATGTAGTACTGTACGTACAGCCGGCTGGACCTCCAGC

ThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeuAlaAlaTyrCysLeuSer
2821 TCACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTTTGGCCGCGTATTGCCTGT
AGTGCTCGTGGACCCACGAGCAACCGCCGCAGGACCGACGAAACCGGCGCATAACGGACA

ThrGlyCysValValIleValGlyArgValValLeuSerGlyLysProAlaIleIlePro

FIG. 26-3

```
2881 CAACAGGCTGCGTGGTCATAGTGGGCAGGGTCGTCTTGTCCGGGAAGCCGGCAATCATAC
     GTTGTCCGACGCACCAGTATCACCCGTCCCAGCAGAACAGGCCCTTCGGCCGTTAGTATG

      AspArgGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuPro
2941 CTGACAGGGAAGTCCTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTAC
     GACTGTCCCTTCAGGAGATGGCTCTCAAGCTACTCTACCTTCTCACGAGAGTCGTGAATG

      TyrIleGluGlnGlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeuLeu
3001 CGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCC
     GCATGTAGCTCGTTCCCTACTACGAGCGGCTCGTCAAGTTCGTCTTCCGGGAGCCGGAGG

      GlnThrAlaSerArgGlnAlaGluValIleAlaProAlaValGlnThrAsnTrpGlnLys
3061 TGCAGACCGCGTCCCGTCAGGCAGAGGTTATCGCCCCTGCTGTCCAGACCAACTGGCAAA
     ACGTCTGGCGCAGGGCAGTCCGTCTCCAATAGCGGGGACGACAGGTCTGGTTGACCGTTT

      LeuGluThrPheTrpAlaLysHisMetTrpAsnPheIleSerGlyIleGlnTyrLeuAla
3121 AACTCGAGACCTTCTGGGCGAAGCATATGTGGAACTTCATCAGTGGGATACAATACTTGG
     TTGAGCTCTGGAAGACCCGCTTCGTATACACCTTGAAGTAGTCACCCTATGTTATGAACC

      GlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeuMetAlaPheThrAlaAla
3181 CGGGCTTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACAGCTG
     GCCCGAACAGTTGCGACGGACCATTGGGGCGGTAACGAAGTAACTACCGAAAATGTCGAC

      ValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsnIleLeuGlyGlyTrpVal
3241 CTGTCACCAGCCCACTAACCACTAGCCAAACCCTCCTCTTCAACATATTGGGGGGGTGGG
     GACAGTGGTCGGGTGATTGGTGATCGGTTTGGGAGGAGAAGTTGTATAACCCCCCCACCC

      AlaAlaGlnLeuAlaAlaProGlyAlaAlaThrAlaPheValGlyAlaGlyLeuAlaGly
3301 TGGCTGCCCAGCTCGCCGCCCCCGGTGCCGCTACTGCCTTTGTGGGCGCTGGCTTAGCTG
     ACCGACGGGTCGAGCGGCGGGGGCCACGGCGATGACGGAAACACCCGCGACCGAATCGAC

      AlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAspIleLeuAlaGlyTyrGly
3361 GCGCCGCCATCGGCAGTGTTGGACTGGGGAAGGTCCTCATAGACATCCTTGCAGGGTATG
     CGCGGCGGTAGCCGTCACAACCTGACCCCTTCCAGGAGTATCTGTAGGAACGTCCCATAC

      AlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSerGlyGluValProSerThr
3421 GCGCGGGCGTGGCGGGAGCTCTTGTGGCATTCAAGATCATGAGCGGTGAGGTCCCCTCCA
     CGCGCCCGCACCGCCCTCGAGAACACCGTAAGTTCTAGTACTCGCCACTCCAGGGGAGGT

      GluAspLeuValAsnLeuLeuProAlaIleLeuSerProGlyAlaLeuValValGlyVal
3481 CGGAGGACCTGGTCAATCTACTGCCCGCCATCCTCTCGCCCGGAGCCCTCGTAGTCGGCG
     GCCTCCTGGACCAGTTAGATGACGGGCGGTAGGAGAGCGGGCCTCGGGAGCATCAGCCGC

      ValCysAlaAlaIleLeuArgArgHisValGlyProGlyGluGlyAlaValGlnTrpMet
3541 TGGTCTGTGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAGGGGGCAGTGCAGTGGA
     ACCAGACACGTCGTTATGACGCGGCCGTGCAACCGGGCCCGCTCCCCCGTCACGTCACCT

      AsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSerProThrHisTyrValPro
3601 TGAACCGGCTGATAGCCTTCGCCTCCCGGGGGAACCATGTTTCCCCCACGCACTACGTGC
     ACTTGGCCGACTATCGGAAGCGGAGGGCCCCCTTGGTACAAAGGGGGTGCGTGATGCACG

      GluSerAspAlaAlaAlaArgValThrAlaIleLeuSerSerLeuThrValThrGlnLeu
3661 CGGAGAGCGATGCAGCTGCCCGCGTCACTGCCATACTCAGCAGCCTCACTGTAACCCAGC
     GCCTCTCGCTACGTCGACGGGCGCAGTGACGGTATGAGTCGTCGGAGTGACATTGGGTCG

      LeuArgArgLeuHisGlnTrpIleSerSerGluCysThrThrProCysSerGlySerTrp
3721 TCCTGAGGCGACTGCACCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGGTTCCT
     AGGACTCCGCTGACGTGGTCACCTATTCGAGCCTCACATGGTGAGGTACGAGGCCAAGGA

      LeuArgAspIleTrpAspTrpIleCysGluValLeuSerAspPheLysThrTrpLeuLys
3781 GGCTAAGGGACATCTGGGACTGGATATGCGAGGTGTTGAGCGACTTTAAGACCTGGCTAA
     CCGATTCCCTGTAGACCCTGACCTATACGCTCCACAACTCGCTGAAATTCTGGACCGATT

      AlaLysLeuMetProGlnLeuProGlyIleProPheValSerCysGlnArgGlyTyrLys
3841 AAGCTAAGCTCATGCCACAGCTGCCTGGGATCCCCTTTGTGTCCTGCCAGCGCGGGTATA
     TTCGATTCGAGTACGGTGTCGACGGACCCTAGGGGAAACACAGGACGGTCGCGCCCATAT
```

FIG. 26-4

GlyValTrpArgValAspGlyIleMetHisThrArgCysHisCysGlyAlaGluIleThr
3901 AGGGGGTCTGGCGAGTGGACGGCATCATGCACACTCGCTGCCACTGTGGAGCTGAGATCA
TCCCCCAGACCGCTCACCTGCCGTAGTACGTGTGAGCGACGGTGACACCTCGACTCTAGT

GlyHisValLysAsnGlyThrMetArgIleValGlyProArgThrCysArgAsnMetTrp
3961 CTGGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTAGGACCTGCAGGAACATGT
GACCTGTACAGTTTTTGCCCTGCTACTCCTAGCAGCCAGGATCCTGGACGTCCTTGTACA

SerGlyThrPheProIleAsnAlaTyrThrThrGlyProCysThrProLeuProAlaPro
4021 GGAGTGGGACCTTCCCCATTAATGCCTACACCACGGGCCCCTGTACCCCCCTTCCTGCGC
CCTCACCCTGGAAGGGGTAATTACGGATGTGGTGCCCGGGGACATGGGGGGAAGGACGCG

AsnTyrThrPheAlaLeuTrpArgValSerAlaGluGluTyrValGluIleArgGlnVal
4081 CGAACTACACGTTCGCGCTATGGAGGGTGTCTGCAGAGGAATATGTGGAGATAAGGCAGG
GCTTGATGTGCAAGCGCGATACCTCCCACAGACGTCTCCTTATACACCTCTATTCCGTCC

GlyAspPheHisTyrValThrGlyMetThrThrAspAsnLeuLysCysProCysGlnVal
4141 TGGGGGACTTCCACTACGTGACGGGTATGACTACTGACAATCTCAAATGCCCGTGCCAGG
ACCCCCTGAAGGTGATGCACTGCCCATACTGATGACTGTTAGAGTTTACGGGCACGGTCC

ProSerProGluPhePheThrGluLeuAspGlyValArgLeuHisArgPheAlaProPro
4201 TCCCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCCTACATAGGTTTGCGCCCC
AGGGTAGCGGGCTTAAAAAGTGTCTTAACCTGCCCCACGCGGATGTATCCAAACGCGGGG

CysLysProLeuLeuArgGluGluValSerPheArgValGlyLeuHisGluTyrProVal
4261 CCTGCAAGCCCTTGCTGCGGGAGGAGGTATCATTCAGAGTAGGACTCCACGAATACCCGG
GGACGTTCGGGAACGACGCCCTCCTCCATAGTAAGTCTCATCCTGAGGTGCTTATGGGCC

GlySerGlnLeuProCysGluProGluProAspValAlaValLeuThrSerMetLeuThr
4321 TAGGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTGGCCGTGTTGACGTCCATGCTCA
ATCCCAGCGTTAATGGAACGCTCGGGCTTGGCCTGCACCGGCACAACTGCAGGTACGAGT

AspProSerHisIleThrAlaGluAlaAlaGlyArgArgLeuAlaArgGlySerProPro
4381 CTGATCCCTCCCATATAACAGCAGAGGCGGCCGGGCGAAGGTTGGCGAGGGGATCACCCC
GACTAGGGAGGGTATATTGTCGTCTCCGCCGGCCCGCTTCCAACCGCTCCCCTAGTGGGG

SerValAlaSerSerSerAlaSerGlnLeuSerAlaProSerLeuLysAlaThrCysThr
4441 CCTCTGTGGCCAGCTCCTCGGCTAGCCAGCTATCCGCTCCATCTCTCAAGGCAACTTGCA
GGAGACACCGGTCGAGGAGCCGATCGGTCGATAGGCGAGGTAGAGAGTTCCGTTGAACGT

AlaAsnHisAspSerProAspAlaGluLeuIleGluAlaAsnLeuLeuTrpArgGlnGlu
4501 CCGCTAACCATGACTCCCCTGATGCTGAGCTCATAGAGGCCAACCTCCTATGGAGGCAGG
GGCGATTGGTACTGAGGGGACTACGACTCGAGTATCTCCGGTTGGAGGATACCTCCGTCC

MetGlyGlyAsnIleThrArgValGluSerGluAsnLysValValIleLeuAspSerPhe
4561 AGATGGGCGGCAACATCACCAGGGTTGAGTCAGAAAACAAAGTGGTGATTCTGGACTCCT
TCTACCCGCCGTTGTAGTGGTCCCAACTCAGTCTTTTGTTTCACCACTAAGACCTGAGGA

AspProLeuValAlaGluGluAspGluArgGluIleSerValProAlaGluIleLeuArg
4621 TCGATCCGCTTGTGGCGGAGGAGGACGAGCGGGAGATCTCCGTACCCGCAGAAATCCTGC
AGCTAGGCGAACACCGCCTCCTCCTGCTCGCCCTCTAGAGGCATGGGCGTCTTTAGGACG

LysSerArgArgPheAlaGlnAlaLeuProValTrpAlaArgProAspTyrAsnProPro
4681 GGAAGTCTCGGAGATTCGCCCAGGCCCTGCCCGTTTGGGCGCGGCCGGACTATAACCCCC
CCTTCAGAGCCTCTAAGCGGGTCCGGGACGGGCAAACCCGCGCCGGCCTGATATTGGGGG

LeuValGluThrTrpLysLysProAspTyrGluProProValValHisGlyCysProLeu
4741 CGCTAGTGGAGACGTGGAAAAAGCCCGACTACGAACCACCTGTGGTCCATGGCTGTCCGC
GCGATCACCTCTGCACCTTTTTCGGGCTGATGCTTGGTGGACACCAGGTACCGACAGGCG

ProProProLysSerProProValProProProArgLysLysArgThrValValLeuThr
4801 TTCCACCTCCAAAGTCCCCTCCTGTGCCTCCGCCTCGGAAGAAGCGGACGGTGGTCCTCA
AAGGTGGAGGTTTCAGGGGAGGACACGGAGGCGGAGCCTTCTTCGCCTGCCACCAGGAGT

GluSerThrLeuSerThrAlaLeuAlaGluLeuAlaThrArgSerPheGlySerSerSer

FIG. 26-5

4861 CTGAATCAACCCTATCTACTGCCTTGGCCGAGCTCGCCACCAGAAGCTTTGGCAGCTCCT
GACTTAGTTGGGATAGATGACGGAACCGGCTCGAGCGGTGGTCTTCGAAACCGTCGAGGA

ThrSerGlyIleThrGlyAspAsnThrThrThrSerSerGluProAlaProSerGlyCys
4921 CAACTTCCGGCATTACGGGCGACAATACGACAACATCCTCTGAGCCCGCCCCTTCTGGCT
GTTGAAGGCCGTAATGCCCGCTGTTATGCTGTTGTAGGAGACTCGGGCGGGGAAGACCGA

ProProAspSerAspAlaGluSerTyrSerSerMetProProLeuGluGlyGluProGly
4981 GCCCCCCCGACTCCGACGCTGAGTCCTATTCCTCCATGCCCCCCCTGGAGGGGGAGCCTG
CGGGGGGGCTGAGGCTGCGACTCAGGATAAGGAGGTACGGGGGGGACCTCCCCCTCGGAC

AspProAspLeuSerAspGlySerTrpSerThrValSerSerGluAlaAsnAlaGluAsp
5041 GGGATCCGGATCTTAGCGACGGGTCATGGTCAACGGTCAGTAGTGAGGCCAACGCGGAGG
CCCTAGGCCTAGAATCGCTGCCCAGTACCAGTTGCCAGTCATCACTCCGGTTGCGCCTCC

ValValCysCysSerMetSerTyrSerTrpThrGlyAlaLeuValThrProCysAlaAla
5101 ATGTCGTGTGCTGCTCAATGTCTTACTCTTGGACAGGCGCACTCGTCACCCCGTGCGCCG
TACAGCACACGACGAGTTACAGAATGAGAACCTGTCCGCGTGAGCAGTGGGGCACGCGGC

GluGluGlnLysLeuProIleAsnAlaLeuSerAsnSerLeuLeuArgHisHisAsnLeu
5161 CGGAAGAACAGAAACTGCCCATCAATGCACTAAGCAACTCGTTGCTACGTCACCACAATT
GCCTTCTTGTCTTTGACGGGTAGTTACGTGATTCGTTGAGCAACGATGCAGTGGTGTTAA

ValTyrSerThrThrSerArgSerAlaCysGlnArgGlnLysLysValThrPheAspArg
5221 TGGTGTATTCCACCACCTCACGCAGTGCTTGCCAAAGGCAGAAGAAAGTCACATTTGACA
ACCACATAAGGTGGTGGAGTGCGTCACGAACGGTTTCCGTCTTCTTTCAGTGTAAACTGT

LeuGlnValLeuAspSerHisTyrGlnAspValLeuLysGluValLysAlaAlaAlaSer
5281 GACTGCAAGTTCTGGACAGCCATTACCAGGACGTACTCAAGGAGGTTAAAGCAGCGGCGT
CTGACGTTCAAGACCTGTCGGTAATGGTCCTGCATGAGTTCCTCCAATTTCGTCGCCGCA

LysValLysAlaAsnLeu
5341 CAAAAGTGAAGGCTAACTTG
GTTTTCACTTCCGATTGAAC

FIG. 26-6

FIG. 27 Translation of DNA 12f

```
      IlePheLysIleArgMetTyrValGlyGlyValGluHisArgLeuGluAlaAlaCysAsn
  1   CCATATTTAAAATCAGGATGTACGTGGGAGGGGTCGAACACAGGCTGGAAGCTGCCTGCA
      GGTATAAATTTTAGTCCTACATGCACCCTCCCCAGCTTGTGTCCGACCTTCGACGGACGT

      TrpThrArgGlyGluArgCysAspLeuGluAspArgAspArgSerGluLeuSerProLeu
 61   ACTGGACGCGGGGCGAACGTTGCGATCTGGAAGACAGGGACAGGTCCGAGCTCAGCCCGT
      TGACCTGCGCCCCGCTTGCAACGCTAGACCTTCTGTCCCTGTCCAGGCTCGAGTCGGGCA

      LeuLeuThrThrThrGlnTrpGlnValLeuProCysSerPheThrThrLeuProAlaLeu
121   TACTGCTGACCACTACACAGTGGCAGGTCCTCCCGTGTTCCTTCACAACCCTACCAGCCT
      ATGACGACTGGTGATGTGTCACCGTCCAGGAGGGCACAAGGAAGTGTTGGGATGGTCGGA

      SerThrGlyLeuIleHisLeuHisGlnAsnIleValAspValGlnTyrLeuTyrGlyVal
181   TGTCCACCGGCCTCATCCACCTCCACCAGAACATTGTGGACGTGCAGTACTTGTACGGGG
      ACAGGTGGCCGGAGTAGGTGGAGGTGGTCTTGTAACACCTGCACGTCATGAACATGCCCC
                                                ---------------------

      GlySerSerIleAlaSerTrpAlaIleLysTrpGluTyrValValLeuLeuPheLeuLeu
241   TGGGGTCAAGCATCGCGTCCTGGGCCATTAAGTGGGAGTACGTCGTTCTCCTGTTCCTTC
      ACCCCAGTTCGTAGCGCAGGACCCGGTAATTCACCCTCATGCAGCAAGAGGACAAGGAAG
      ------------------------------------------------------------

      LeuAlaAspAlaArgValCysSerCysLeuTrpMetMetLeuLeuIleSerGlnAlaGlu
301   TGCTTGCAGACGCGCGCGTCTGCTCCTGCTTGTGGATGATGCTACTCATATCCCAAGCGG
      ACGAACGTCTGCGCGCGCAGACGAGGACGAACACCTACTACGATGAGTATAGGGTTCGCC
      ------------------Overlap with 14i-------------------------

      AlaAlaLeuGluAsnLeuValIleLeuAsnAlaAlaSerLeuAlaGlyThrHisGlyLeu
361   AGGCGGCTTTGGAGAACCTCGTAATACTTAATGCAGCATCCCTGGCCGGGACGCACGGTC
      TCCGCCGAAACCTCTTGGAGCATTATGAATTACGTCGTAGGGACCGGCCCTGCGTGCCAG
      -------

      Val
421   TTGTATC
      AACATAG
```

FIG. 28 Translation of DNA 35f

```
    ----------Overlap with 39c--------------------
      LeuLysGluValLysAlaAlaAlaSerLysValLysAlaAsnLeuLeuSerValGluGlu
  1   TGCTCAAGGAGGTTAAAGCAGCGGCGTCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGG
      ACGAGTTCCTCCAATTTCGTCGCCGCAGTTTTCACTTCCGATTGAACGATAGGCATCTCC

      AlaCysSerLeuThrProProHisSerAlaLysSerLysPheGlyTyrGlyAlaLysAsp
 61   AAGCTTGCAGCCTGACGCCCCCACACTCAGCCAAATCCAAGTTTGGTTATGGGGCAAAAG
      TTCGAACGTCGGACTGCGGGGGTGTGAGTCGGTTTAGGTTCAAACCAATACCCCGTTTTC

      ValArgCysHisAlaArgLysAlaValThrHisIleAsnSerValTrpLysAspLeuLeu
121   ACGTCCGTTGCCATGCCAGAAAGGCCGTAACCCACATCAACTCCGTGTGGAAAGACCTTC
      TGCAGGCAACGGTACGGTCTTTCCGGCATTGGGTGTAGTTGAGGCACACCTTTCTGGAAG

      GluAspAsnValThrProIleAspThrThrIleMetAlaLysAsnGluValPheCysVal
181   TGGAAGACAATGTAACACCAATAGACACTACCATCATGGCTAAGAACGAGGTTTTCTGCG
      ACCTTCTGTTACATTGTGGTTATCTGTGATGGTAGTACCGATTCTTGCTCCAAAAGACGC

      GlnProGluLysGlyGlyArgLysProAlaArgLeuIleValPheProAspLeuGlyVal
241   TTCAGCCTGAGAAGGGGGGTCGTAAGCCAGCTCGTCTCATCGTGTTCCCCGATCTGGGCG
      AAGTCGGACTCTTCCCCCCAGCATTCGGTCGAGCAGAGTAGCACAAGGGGCTAGACCCGC

      ArgValCysGluLysMetAlaLeuTyrAspValValThrLysLeuProLeuAlaValMet
301   TGCGCGTGTGCGAAAAGATGGCTTTGTACGACGTGGTTACAAAGCTCCCCTTGGCCGTGA
      ACGCGCACACGCTTTTCTACCGAAACATGCTGCACCAATGTTTCGAGGGGAACCGGCACT

      GlySerSerTyrGlyPheGlnTyrSerProGlyGlnArgValGluPheLeuValGlnAla
361   TGGGAAGCTCCTACGGATTCCAATACTCACCAGGACAGCGGGTTGAATTCCTCGTGCAAG
      ACCCTTCGAGGATGCCTAAGGTTATGAGTGGTCCTGTCGCCCAACTTAAGGAGCACGTTC

      TrpLysSerLysLysThrProMetGlyPheSerTyrAspThrArgCysPheAspSerThr
421   CGTGGAAGTCCAAGAAAACCCCAATGGGGTTCTCGTATGATACCCGCTGCTTTGACTCCA
      GCACCTTCAGGTTCTTTTGGGGTTACCCCAAGAGCATACTATGGGCGACGAAACTGAGGT

      ValThrGluSerAspIleArgThrGluGluAla
481   CAGTCACTGAGAGCGACATCCGTACGGAGGAGGCA
      GTCAGTGACTCTCGCTGTAGGCATGCCTCCTCCGT
```

FIG. 29 Translation of DNA 19g

```
     -------------------------------------------------------------
     GluPheLeuValGlnAlaTrpLysSerLysLysThrProMetGlyPheSerTyrAspThr
  1  GAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAAACCCCAATGGGGTTCTCGTATGATACC
     CTTAAGGAGCACGTTCGCACCTTCAGGTTCTTTTGGGGTTACCCCAAGAGCATACTATGG

     -------------Overlap with 35f----------------------
     ArgCysPheAspSerThrValThrGluSerAspIleArgThrGluGluAlaIleTyrGln
 61  CGCTGCTTTGACTCCACAGTCACTGAGAGCGACATCCGTACGGAGGAGGCAATCTACCAA
     GCGACGAAACTGAGGTGTCAGTGACTCTCGCTGTAGGCATGCCTCCTCCGTTAGATGGTT

     CysCysAspLeuAspProGlnAlaArgValAlaIleLysSerLeuThrGluArgLeuTyr
121  TGTTGTGACCTCGACCCCCAAGCCCGCGTGGCCATCAAGTCCCTCACCGAGAGGCTTTAT
     ACAACACTGGAGCTGGGGGTTCGGGCGCACCGGTAGTTCAGGGAGTGGCTCTCCGAAATA

     ValGlyGlyProLeuThrAsnSerArgGlyGluAsnCysGlyTyrArgArgCysArgAla
181  GTTGGGGGCCCTCTTACCAATTCAAGGGGGGAGAACTGCGGCTATCGCAGGTGCCGCGCG
     CAACCCCCGGGAGAATGGTTAAGTTCCCCCCTCTTGACGCCGATAGCGTCCACGGCGCGC

     SerGlyValLeuThrThrSerCysGlyAsnThrLeuThrCysTyrIleLysAlaArgAla
241  AGCGGCGTACTGACAACTAGCTGTGGTAACACCCTCACTTGCTACATCAAGGCCCGGGCA
     TCGCCGCATGACTGTTGATCGACACCATTGTGGGAGTGAACGATGTAGTTCCGGGCCCGT

     AlaCysArgAlaAlaGlyLeuGlnAspCysThrMetLeuValCysGlyAspAspLeuVal
301  GCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTCGTGTGTGGCGACGACTTAGTC
     CGGACAGCTCGGCGTCCCGAGGTCCTGACGTGGTACGAGCACACACCGCTGCTGAATCAG

     ValIleCysGluSerAlaGlyValGlnGluAspAlaAla
361  GTTATCTGTGAAAGCGCGGGGGTCCAGGAGGACGCGGCGAG
     CAATAGACACTTTCGCGCCCCCAGGTCCTCCTGCGCCGCTC
```

FIG. 30 Translation of DNA 26g

```
     ------------------------------------------------------------
     GlyGlyGluAsnCysGlyTyrArgArgCysArgAlaSerGlyValLeuThrThrSerCys
  1  GGGGGGGAGAACTGCGGCTATCGCAGGTGCCGCGCAAGCGGCGTACTGACAACTAGCTGT
     CCCCCCCTCTTGACGCCGATAGCGTCCACGGCGCGTTCGCCGCATGACTGTTGATCGACA

     ------------------------------------------------------------
     GlyAsnThrLeuThrCysTyrIleLysAlaArgAlaAlaCysArgAlaAlaGlyLeuGln
 61  GGTAACACCCTCACTTGTTACATCAAGGCCCGAGCAGCCTGTCGAGCCGCAGGGCTCCAG
     CCATTGTGGGAGTGAACAATGTAGTTCCGGGCTCGTCGGACAGCTCGGCGTCCCGAGGTC

     ------------Overlap with 19g--------------------------------
     AspCysThrMetLeuValCysGlyAspAspLeuValValIleCysGluSerAlaGlyVal
121  GACTGCACCATGCTCGTGTGTGGCGACGACTTAGTCGTTATCTGTGAAAGCGCGGGGTC
     CTGACGTGGTACGAGCACACACCGCTGCTGAATCAGCAATAGACACTTTCGCGCCCCCAG

     -----------------
     GlnGluAspAlaAlaSerLeuArgAlaPheThrGluAlaMetThrArgTyrSerAlaPro
181  CAGGAGGACGCGGCGAGCCTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCC
     GTCCTCCTGCGCCGCTCGGACTCTCGGAAGTGCCTCCGATACTGGTCCATGAGGCGGGGG

     ProGlyAspProProGlnProGluTyrAspLeuGluLeuIleThrSerCysSerSerAsn
241  CCTGGGGACCCCCCACAACCAGAATACGACTTGGAGCTCATAACATCATGCTCCTCCAAC
     GGACCCCTGGGGGGTGTTGGTCTTATGCTGAACCTCGAGTATTGTAGTACGAGGAGGTTG

     ValSerValAlaHisAspGlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThr
301  GTGTCAGTCGCCCACGACGGCGCTGGAAAGAGGGTCTACTACCTCACCCGTGACCCTACA
     CACAGTCAGCGGGTGCTGCCGCGACCTTTCTCCCAGATGATGGAGTGGGCACTGGGATGT

     ThrProLeuAlaArgAlaAlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeu
361  ACCCCCCTCGCGAGAGCTGCGTGGGAGACAGCAAGACACACTCCAGTCAATTCCTGGCTA
     TGGGGGGAGCGCTCTCGACGCACCCTCTGTCGTTCTGTGTGAGGTCAGTTAAGGACCGAT

     GlyAsnIleIleMetPheAlaProThrLeuTrpAla
421  GGCAACATAATCATGTTTGCCCCCACACTGTGGGCG
     CCGTTGTATTAGTACAAACGGGGGTGTGACACCCGC
```

FIG. 31 Translation of DNA 15e

```
     ------------------------------------------------------------
     GlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThrThrProLeuAlaArgAla
  1  CGGCGCTGGAAAGAGGGTCTACTACCTCACCCGTGACCCTACAACCCCCCTCGCGAGAGC
     GCCGCGACCTTTCTCCCAGATGATGGAGTGGGCACTGGGATGTTGGGGGGAGCGCTCTCG

     ------------------Overlap with 26g--------------------------
     AlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeuGlyAsnIleIleMetPhe
 61  TGCGTGGGAGACAGCAAGACACACTCCAGTCAATTCCTGGCTAGGCAACATAATCATGTT
     ACGCACCCTCTGTCGTTCTGTGTGAGGTCAGTTAAGGACCGATCCGTTGTATTAGTACAA

     -------------------
     AlaProThrLeuTrpAlaArgMetIleLeuMetThrHisPhePheSerValLeuIleAla
121  TGCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTTATAGC
     ACGGGGGTGTGACACCCGCTCCTACTATGACTACTGGGTAAAGAAATCGCAGGAATATCG

     ArgAspGlnLeuGluGlnAlaLeuAspCysGluIleTyrGlyAlaCysTyrSerIleGlu
181  CAGGGACCAGCTTGAACAGGCCCTCGATTGCGAGATCTACGGGGCCTGCTACTCCATAGA
     GTCCCTGGTCGAACTTGTCCGGGAGCTAACGCTCTAGATGCCCCGGACGATGAGGTATCT

     ProLeuAspLeuProProIleIleGlnArgLeu
241  ACCACTTGATCTACCTCCAATCATTCAAAGACTC
     TGGTGAACTAGATGGAGGTTAGTAAGTTTCTGAG
```

FIG. 32-1 COMBINED ORF OF DNAs 12f through 15e

```
     IlePheLysIleArgMetTyrValGlyGlyValGluHisArgLeuGluAlaAlaCysAsn
1    CCATATTTAAAATCAGGATGTACGTGGGAGGGGTCGAACACAGGCTGGAAGCTGCCTGCA
     GGTATAAATTTTAGTCCTACATGCACCCTCCCCAGCTTGTGTCCGACCTTCGACGGACGT

     TrpThrArgGlyGluArgCysAspLeuGluAspArgAspArgSerGluLeuSerProLeu
61   ACTGGACGCGGGGCGAACGTTGCGATCTGGAAGACAGGGACAGGTCCGAGCTCAGCCCGT
     TGACCTGCGCCCCGCTTGCAACGCTAGACCTTCTGTCCCTGTCCAGGCTCGAGTCGGGCA

     LeuLeuThrThrThrGlnTrpGlnValLeuProCysSerPheThrThrLeuProAlaLeu
121  TACTGCTGACCACTACACAGTGGCAGGTCCTCCCGTGTTCCTTCACAACCCTACCAGCCT
     ATGACGACTGGTGATGTGTCACCGTCCAGGAGGGCACAAGGAAGTGTTGGGATGGTCGGA

     SerThrGlyLeuIleHisLeuHisGlnAsnIleValAspValGlnTyrLeuTyrGlyVal
181  TGTCCACCGGCCTCATCCACCTCCACCAGAACATTGTGGACGTGCAGTACTTGTACGGGG
     ACAGGTGGCCGGAGTAGGTGGAGGTGGTCTTGTAACACCTGCACGTCATGAACATGCCCC

     GlySerSerIleAlaSerTrpAlaIleLysTrpGluTyrValValLeuLeuPheLeuLeu
241  TGGGGTCAAGCATCGCGTCCTGGGCCATTAAGTGGGAGTACGTCGTTCTCCTGTTCCTTC
     ACCCCAGTTCGTAGCGCAGGACCCGGTAATTCACCCTCATGCAGCAAGAGGACAAGGAAG

     LeuAlaAspAlaArgValCysSerCysLeuTrpMetMetLeuLeuIleSerGlnAlaGlu
301  TGCTTGCAGACGCGCGCGTCTGCTCCTGCTTGTGGATGATGCTACTCATATCCCAAGCGG
     ACGAACGTCTGCGCGCGCAGACGAGGACGAACACCTACTACGATGAGTATAGGGTTCGCC

     AlaAlaLeuGluAsnLeuValIleLeuAsnAlaAlaSerLeuAlaGlyThrHisGlyLeu
361  AGGCGGCTTTGGAGAACCTCGTAATACTTAATGCAGCATCCCTGGCCGGGACGCACGGTC
     TCCGCCGAAACCTCTTGGAGCATTATGAATTACGTCGTAGGGACCGGCCCTGCGTGCCAG

     ValSerPheLeuValPhePheCysPheAlaTrpTyrLeuLysGlyLysTrpValProGly
421  TTGTATCCTTCCTCGTGTTCTTCTGCTTTGCATGGTATTTGAAGGGTAAGTGGGTGCCCG
     AACATAGGAAGGAGCACAAGAAGACGAAACGTACCATAAACTTCCCATTCACCCACGGGC

     AlaValTyrThrPheTyrGlyMetTrpProLeuLeuLeuLeuLeuLeuAlaLeuProGln
481  GAGCGGTCTACACCTTCTACGGGATGTGGCCTCTCCTCCTGCTCCTGTTGGCGTTGCCCC
     CTCGCCAGATGTGGAAGATGCCCTACACCGGAGAGGAGGACGAGGACAACCGCAACGGGG

     ArgAlaTyrAlaLeuAspThrGluValAlaAlaSerCysGlyGlyValValLeuValGly
541  AGCGGGCGTACGCGCTGGACACGGAGGTGGCCGCGTCGTGTGGCGGTGTTGTTCTCGTCG
     TCGCCCGCATGCGCGACCTGTGCCTCCACCGGCGCAGCACACCGCCACAACAAGAGCAGC

     LeuMetAlaLeuThrLeuSerProTyrTyrLysArgTyrIleSerTrpCysLeuTrpTrp
601  GGTTGATGGCGCTGACTCTGTCACCATATTACAAGCGCTATATCAGCTGGTGCTTGTGGT
     CCAACTACCGCGACTGAGACAGTGGTATAATGTTCGCGATATAGTCGACCACGAACACCA

     LeuGlnTyrPheLeuThrArgValGluAlaGlnLeuHisValTrpIleProProLeuAsn
661  GGCTTCAGTATTTTCTGACCAGAGTGGAAGCGCAACTGCACGTGTGGATTCCCCCCCTCA
     CCGAAGTCATAAAAGACTGGTCTCACCTTCGCGTTGACGTGCACACCTAAGGGGGGGAGT

     ValArgGlyGlyArgAspAlaValIleLeuLeuMetCysAlaValHisProThrLeuVal
721  ACGTCCGAGGGGGGCGCGACGCCGTCATCTTACTCATGTGTGCTGTACACCCGACTCTGG
     TGCAGGCTCCCCCCGCGCTGCGGCAGTAGAATGAGTACACACGACATGTGGGCTGAGACC

     PheAspIleThrLysLeuLeuLeuAlaValPheGlyProLeuTrpIleLeuGlnAlaSer
781  TATTTGACATCACCAAATTGCTGCTGGCCGTCTTCGGACCCCTTTGGATTCTTCAAGCCA
     ATAAACTGTAGTGGTTTAACGACGACCGGCAGAAGCCTGGGGAAACCTAAGAAGTTCGGT

     LeuLeuLysValProTyrPheValArgValGlnGlyLeuLeuArgPheCysAlaLeuAla
841  GTTTGCTTAAAGTACCCTACTTTGTGCGCGTCCAAGGCCTTCTCCGGTTCTGCGCGTTAG
     CAAACGAATTTCATGGGATGAAACACGCGCAGGTTCCGGAAGAGGCCAAGACGCGCAATC

     ArgLysMetIleGlyGlyHisTyrValGlnMetValIleIleLysLeuGlyAlaLeuThr
901  CGCGGAAGATGATCGGAGGCCATTACGTGCAAATGGTCATCATTAAGTTAGGGGCGCTTA
     GCGCCTTCTACTAGCCTCCGGTAATGCACGTTTACCAGTAGTAATTCAATCCCCGCGAAT
```

33/63

```
      GlyThrTyrValTyrAsnHisLeuThrProLeuArgAspTrpAlaHisAsnGlyLeuArg
 961  CTGGCACCTATGTTTATAACCATCTCACTCCTCTTCGGGACTGGGCGCACAACGGCTTGC
      GACCGTGGATACAAATATTGGTAGAGTGAGGAGAAGCCCTGACCCGCGTGTTGCCGAACG

      AspLeuAlaValAlaValGluProValValPheSerGlnMetGluThrLysLeuIleThr
1021  GAGATCTGGCCGTGGCTGTAGAGCCAGTCGTCTTCTCCCAAATGGAGACCAAGCTCATCA
      CTCTAGACCGGCACCGACATCTCGGTCAGCAGAAGAGGGTTTACCTCTGGTTCGAGTAGT

      TrpGlyAlaAspThrAlaAlaCysGlyAspIleIleAsnGlyLeuProValSerAlaArg
1081  CGTGGGGGGCAGATACCGCCGCGTGCGGTGACATCATCAACGGCTTGCCTGTTTCCGCCC
      GCACCCCCCGTCTATGGCGGCGCACGCCACTGTAGTAGTTGCCGAACGGACAAAGGCGGG

      ArgGlyArgGluIleLeuLeuGlyProAlaAspGlyMetValSerLysGlyTrpArgLeu
1141  GCAGGGGCCGGGAGATACTGCTCGGGCCAGCCGATGGAATGGTCTCCAAGGGGTGGAGGT
      CGTCCCCGGCCCTCTATGACGAGCCCGGTCGGCTACCTTACCAGAGGTTCCCCACCTCCA

      LeuAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr
1201  TGCTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCA
      ACGACCGCGGGTAGTGCCGCATGCGGGTCGTCTGTTCCCCGGAGGATCCCACGTATTAGT

      SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla
1261  CCAGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTG
      GGTCGGATTGACCGGCCCTGTTTTTGGTTCACCTCCCACTCCAGGTCTAACACAGTTGAC

      AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla
1321  CTGCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGG
      GACGGGTTTGGAAGGACCGTTGCACGTAGTTACCCCACACGACCTGACAGATGGTGCCCC

      GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp
1381  CCGGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAG
      GGCCTTGCTCCTGGTAGCGCAGTGGGTTCCCAGGACAGTAGGTCTACATATGGTTACATC

      GlnAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys
1441  ACCAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTT
      TGGTTCTGGAACACCCGACCGGGCGAGGCGTTCCATCGGCGAGTAACTGTGGGACGTGAA

      GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg
1501  GCGGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGC
      CGCCGAGGAGCCTGGAAATGGACCAGTGCTCCGTGCGGCTACAGTAAGGGCACGCGGCCG

      GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer
1561  GGGGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCT
      CCCCACTATCGTCCCCGTCGGACGACAGCGGGGCCGGGTAAAGGATGAACTTTCCGAGGA

      GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys
1621  CGGGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGT
      GCCCCCCAGGCGACAACACGGGGCGCCCCGTGCGGCACCCGTATAAATCCCGGCGCCACA

      ThrArgGlyValAlaLysAlaValAspPheIleProValGluAsnLeuGluThrThrMet
1681  GCACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCA
      CGTGGGCACCTCACCGATTCCGCCACCTGAAATAGGGACACCTCTTGGATCTCTGTTGGT

      ArgSerProValPheThrAspAsnSerSerProProValValProGlnSerPheGlnVal
1741  TGAGGTCCCCGGTGTTCACGGATAACTCCTCTCCACCAGTAGTGCCCCAGAGCTTCCAGG
      ACTCCAGGGGCCACAAGTGCCTATTGAGGAGAGGTGGTCATCACGGGGTCTCGAAGGTCC

      AlaHisLeuHisAlaProThrGlySerGlyLysSerThrLysValProAlaAlaTyrAla
1801  TGGCTCACCTCCATGCTCCCACAGGCAGCGGCAAAAGCACCAAGGTCCCGGCTGCATATG
      ACCGAGTGGAGGTACGAGGGTGTCCGTCGCCGTTTTCGTGGTTCCAGGGCCGACGTATAC

      AlaGlnGlyTyrLysValLeuValLeuAsnProSerValAlaAlaThrLeuGlyPheGly
1861  CAGCTCAGGGCTATAAGGTGCTAGTACTCAACCCCTCTGTTGCTGCAACACTGGGCTTTG
      GTCGAGTCCCGATATTCCACGATCATGAGTTGGGGAGACAACGACGTTGTGACCCGAAAC

      AlaTyrMetSerLysAlaHisGlyIleAspProAsnIleArgThrGlyValArgThrIle
```

FIG. 32-2

1921 GTGCTTACATGTCCAAGGCTCATGGGATCGATCCTAACATCAGGACCGGGGTGAGAACAA
CACGAATGTACAGGTTCCGAGTACCCTAGCTAGGATTGTAGTCCTGGCCCCACTCTTGTT

ThrThrGlySerProIleThrTyrSerThrTyrGlyLysPheLeuAlaAspGlyGlyCys
1981 TTACCACTGGCAGCCCCATCACGTACTCCACCTACGGCAAGTTCCTTGCCGACGGCGGGT
AATGGTGACCGTCGGGGTAGTGCATGAGGTGGATGCCGTTCAAGGAACGGCTGCCGCCCA

SerGlyGlyAlaTyrAspIleIleIleCysAspGluCysHisSerThrAspAlaThrSer
2041 GCTCGGGGGGCGCTTATGACATAATAATTTGTGACGAGTGCCACTCCACGGATGCCACAT
CGAGCCCCCCGCGAATACTGTATTATTAAACACTGCTCACGGTGAGGTGCCTACGGTGTA

IleLeuGlyIleGlyThrValLeuAspGlnAlaGluThrAlaGlyAlaArgLeuValVal
2101 CCATCTTGGGCATCGGCACTGTCCTTGACCAAGCAGAGACTGCGGGGGCGAGACTGGTTG
GGTAGAACCCGTAGCCGTGACAGGAACTGGTTCGTCTCTGACGCCCCCGCTCTGACCAAC

LeuAlaThrAlaThrProProGlySerValThrValProHisProAsnIleGluGluVal
2161 TGCTCGCCACCGCCACCCCTCCGGGCTCCGTCACTGTGCCCCATCCCAACATCGAGGAGG
ACGAGCGGTGGCGGTGGGGAGGCCCGAGGCAGTGACACGGGGTAGGGTTGTAGCTCCTCC

AlaLeuSerThrThrGlyGluIleProPheTyrGlyLysAlaIleProLeuGluValIle
2221 TTGCTCTGTCCACCACCGGAGAGATCCCTTTTTACGGCAAGGCTATCCCCCTCGAAGTAA
AACGAGACAGGTGGTGGCCTCTCTAGGGAAAAATGCCGTTCCGATAGGGGGAGCTTCATT

LysGlyGlyArgHisLeuIlePheCysHisSerLysLysLysCysAspGluLeuAlaAla
2281 TCAAGGGGGGGAGACATCTCATCTTCTGTCATTCAAAGAAGAAGTGCGACGAACTCGCCG
AGTTCCCCCCCTCTGTAGAGTAGAAGACAGTAAGTTTCTTCTTCACGCTGCTTGAGCGGC

LysLeuValAlaLeuGlyIleAsnAlaValAlaTyrTyrArgGlyLeuAspValSerVal
2341 CAAAGCTGGTCGCATTGGGCATCAATGCCGTGGCCTACTACCGCGGTCTTGACGTGTCCG
GTTTCGACCAGCGTAACCCGTAGTTACGGCACCGGATGATGGCGCCAGAACTGCACAGGC

IleProThrSerGlyAspValValValValAlaThrAspAlaLeuMetThrGlyTyrThr
2401 TCATCCCGACCAGCGGCGATGTTGTCGTCGTGGCAACCGATGCCCTCATGACCGGCTATA
AGTAGGGCTGGTCGCCGCTACAACAGCAGCACCGTTGGCTACGGGAGTACTGGCCGATAT

GlyAspPheAspSerValIleAspCysAsnThrCysValThrGlnThrValAspPheSer
2461 CCGGCGACTTCGACTCGGTGATAGACTGCAATACGTGTGTCACCCAGACAGTCGATTTCA
GGCCGCTGAAGCTGAGCCACTATCTGACGTTATGCACACAGTGGGTCTGTCAGCTAAAGT

LeuAspProThrPheThrIleGluThrIleThrLeuProGlnAspAlaValSerArgThr
2521 GCCTTGACCCTACCTTCACCATTGAGACAATCACGCTCCCCCAGGATGCTGTCTCCCGCA
CGGAACTGGGATGGAAGTGGTAACTCTGTTAGTGCGAGGGGGTCCTACGACAGAGGGCGT

GlnArgArgGlyArgThrGlyArgGlyLysProGlyIleTyrArgPheValAlaProGly
2581 CTCAACGTCGGGGCAGGACTGGCAGGGGGAAGCCAGGCATCTACAGATTTGTGGCACCGG
GAGTTGCAGCCCCGTCCTGACCGTCCCCCTTCGGTCCGTAGATGTCTAAACACCGTGGCC

GluArgProSerGlyMetPheAspSerSerValLeuCysGluCysTyrAspAlaGlyCys
2641 GGGAGCGCCCCTCCGGCATGTTCGACTCGTCCGTCCTCTGTGAGTGCTATGACGCAGGCT
CCCTCGCGGGGAGGCCGTACAAGCTGAGCAGGCAGGAGACACTCACGATACTGCGTCCGA

AlaTrpTyrGluLeuThrProAlaGluThrThrValArgLeuArgAlaTyrMetAsnThr
2701 GTGCTTGGTATGAGCTCACGCCCGCCGAGACTACAGTTAGGCTACGAGCGTACATGAACA
CACGAACCATACTCGAGTGCGGGCGGCTCTGATGTCAATCCGATGCTCGCATGTACTTGT

ProGlyLeuProValCysGlnAspHisLeuGluPheTrpGluGlyValPheThrGlyLeu
2761 CCCCGGGGCTTCCCGTGTGCCAGGACCATCTTGAATTTTGGGAGGGCGTCTTTACAGGCC
GGGGCCCCGAAGGGCACACGGTCCTGGTAGAACTTAAAACCCTCCCGCAGAAATGTCCGG

ThrHisIleAspAlaHisPheLeuSerGlnThrLysGlnSerGlyGluAsnLeuProTyr
2821 TCACTCATATAGATGCCCACTTTCTATCCCAGACAAAGCAGAGTGGGGAGAACCTTCCTT
AGTGAGTATATCTACGGGTGAAAGATAGGGTCTGTTTCGTCTCACCCCTCTTGGAAGGAA

LeuValAlaTyrGlnAlaThrValCysAlaArgAlaGlnAlaProProProSerTrpAsp
2881 ACCTGGTAGCGTACCAAGCCACCGTGTGCGCTAGGGCTCAAGCCCCTCCCCCATCGTGGG
TGGACCATCGCATGGTTCGGTGGCACACGCGATCCCGAGTTCGGGGAGGGGGTAGCACCC

FIG. 32-3

```
      GlnMetTrpLysCysLeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeu
2941 ACCAGATGTGGAAGTGTTTGATTCGCCTCAAGCCCACCCTCCATGGGCCAACACCCCTGC
     TGGTCTACACCTTCACAAACTAAGCGGAGTTCGGGTGGGAGGTACCCGGTTGTGGGGACG

      TyrArgLeuGlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIle
3001 TATACAGACTGGGCGCTGTTCAGAATGAAATCACCCTGACGCACCCAGTCACCAAATACA
     ATATGTCTGACCCGCGACAAGTCTTACTTTAGTGGGACTGCGTGGGTCAGTGGTTTATGT

      MetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyGly
3061 TCATGACATGCATGTCGGCCGACCTGGAGGTCGTCACGAGCACCTGGGTGCTCGTTGGCG
     AGTACTGTACGTACAGCCGGCTGGACCTCCAGCAGTGCTCGTGGACCCACGAGCAACCGC

      ValLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValValIleValGlyArg
3121 GCGTCCTGGCTGCTTTGGCCGCGTATTGCCTGTCAACAGGCTGCGTGGTCATAGTGGGCA
     CGCAGGACCGACGAAACCGGCGCATAACGGACAGTTGTCCGACGCACCAGTATCACCCGT

      ValValLeuSerGlyLysProAlaIleIleProAspArgGluValLeuTyrArgGluPhe
3181 GGGTCGTCTTGTCCGGGAAGCCGGCAATCATACCTGACAGGGAAGTCCTCTACCGAGAGT
     CCCAGCAGAACAGGCCCTTCGGCCGTTAGTATGGACTGTCCCTTCAGGAGATGGCTCTCA

      AspGluMetGluGluCysSerGlnHisLeuProTyrIleGluGlnGlyMetMetLeuAla
3241 TCGATGAGATGGAAGAGTGCTCTCAGCACTTACCGTACATCGAGCAAGGGATGATGCTCG
     AGCTACTCTACCTTCTCACGAGAGTCGTGAATGGCATGTAGCTCGTTCCCTACTACGAGC

      GluGlnPheLysGlnLysAlaLeuGlyLeuLeuGlnThrAlaSerArgGlnAlaGluVal
3301 CCGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCCTGCAGACCGCGTCCCGTCAGGCAGAGG
     GGCTCGTCAAGTTCGTCTTCCGGGAGCCGGAGGACGTCTGGCGCAGGGCAGTCCGTCTCC

      IleAlaProAlaValGlnThrAsnTrpGlnLysLeuGluThrPheTrpAlaLysHisMet
3361 TTATCGCCCCTGCTGTCCAGACCAACTGGCAAAAACTCGAGACCTTCTGGGCGAAGCATA
     AATAGCGGGGACGACAGGTCTGGTTGACCGTTTTTGAGCTCTGGAAGACCCGCTTCGTAT

      TrpAsnPheIleSerGlyIleGlnTyrLeuAlaGlyLeuSerThrLeuProGlyAsnPro
3421 TGTGGAACTTCATCAGTGGGATACAATACTTGGCGGGCTTGTCAACGCTGCCTGGTAACC
     ACACCTTGAAGTAGTCACCCTATGTTATGAACCGCCCGAACAGTTGCGACGGACCATTGG

      AlaIleAlaSerLeuMetAlaPheThrAlaAlaValThrSerProLeuThrThrSerGln
3481 CCGCCATTGCTTCATTGATGGCTTTTACAGCTGCTGTCACCAGCCCACTAACCACTAGCC
     GGCGGTAACGAAGTAACTACCGAAAATGTCGACGACAGTGGTCGGGTGATTGGTGATCGG

      ThrLeuLeuPheAsnIleLeuGlyGlyTrpValAlaAlaGlnLeuAlaAlaProGlyAla
3541 AAACCCTCCTCTTCAACATATTGGGGGGGTGGGTGGCTGCCCAGCTCGCCGCCCCCGGTG
     TTTGGGAGGAGAAGTTGTATAACCCCCCCACCCACCGACGGGTCGAGCGGCGGGGGCCAC

      AlaThrAlaPheValGlyAlaGlyLeuAlaGlyAlaAlaIleGlySerValGlyLeuGly
3601 CCGCTACTGCCTTTGTGGGCGCTGGCTTAGCTGGCGCCGCCATCGGCAGTGTTGGACTGG
     GGCGATGACGGAAACACCCGCGACCGAATCGACCGCGGCGGTAGCCGTCACAACCTGACC

      LysValLeuIleAspIleLeuAlaGlyTyrGlyAlaGlyValAlaGlyAlaLeuValAla
3661 GGAAGGTCCTCATAGACATCCTTGCAGGGTATGGCGCGGGCGTGGCGGGAGCTCTTGTGG
     CCTTCCAGGAGTATCTGTAGGAACGTCCCATACCGCGCCCGCACCGCCCTCGAGAACACC

      PheLysIleMetSerGlyGluValProSerThrGluAspLeuValAsnLeuLeuProAla
3721 CATTCAAGATCATGAGCGGTGAGGTCCCCTCCACGGAGGACCTGGTCAATCTACTGCCCG
     GTAAGTTCTAGTACTCGCCACTCCAGGGGAGGTGCCTCCTGGACCAGTTAGATGACGGGC

      IleLeuSerProGlyAlaLeuValValGlyValValCysAlaAlaIleLeuArgArgHis
3781 CCATCCTCTCGCCCGGAGCCCTCGTAGTCGGCGTGGTCTGTGCAGCAATACTGCGCCGGC
     GGTAGGAGAGCGGGCCTCGGGAGCATCAGCCGCACCAGACACGTCGTTATGACGCGGCCG

      ValGlyProGlyGluGlyAlaValGlnTrpMetAsnArgLeuIleAlaPheAlaSerArg
3841 ACGTTGGCCCGGGCGAGGGGGCAGTGCAGTGGATGAACCGGCTGATAGCCTTCGCCTCCC
     TGCAACCGGGCCCGCTCCCCCGTCACGTCACCTACTTGGCCGACTATCGGAAGCGGAGGG

      GlyAsnHisValSerProThrHisTyrValProGluSerAspAlaAlaAlaArgValThr
```

FIG. 32-4

3901 GGGGGAACCATGTTTCCCCCACGCACTACGTGCCGGAGAGCGATGCAGCTGCCCGCGTCA
CCCCCTTGGTACAAAGGGGGTGCGTGATGCACGGCCTCTCGCTACGTCGACGGGCGCAGT

AlaIleLeuSerSerLeuThrValThrGlnLeuLeuArgArgLeuHisGlnTrpIleSer
3961 CTGCCATACTCAGCAGCCTCACTGTAACCCAGCTCCTGAGGCGACTGCACCAGTGGATAA
GACGGTATGAGTCGTCGGAGTGACATTGGGTCGAGGACTCCGCTGACGTGGTCACCTATT

SerGluCysThrThrProCysSerGlySerTrpLeuArgAspIleTrpAspTrpIleCys
4021 GCTCGGAGTGTACCACTCCATGCTCCGGTTCCTGGCTAAGGGACATCTGGGACTGGATAT
CGAGCCTCACATGGTGAGGTACGAGGCCAAGGACCGATTCCCTGTAGACCCTGACCTATA

GluValLeuSerAspPheLysThrTrpLeuLysAlaLysLeuMetProGlnLeuProGly
4081 GCGAGGTGTTGAGCGACTTTAAGACCTGGCTAAAAGCTAAGCTCATGCCACAGCTGCCTG
CGCTCCACAACTCGCTGAAATTCTGGACCGATTTTCGATTCGAGTACGGTGTCGACGGAC

IleProPheValSerCysGlnArgGlyTyrLysGlyValTrpArgValAspGlyIleMet
4141 GGATCCCCTTTGTGTCCTGCCAGCGCGGGTATAAGGGGGTCTGGCGAGTGGACGGCATCA
CCTAGGGGAAACACAGGACGGTCGCGCCCATATTCCCCCAGACCGCTCACCTGCCGTAGT

HisThrArgCysHisCysGlyAlaGluIleThrGlyHisValLysAsnGlyThrMetArg
4201 TGCACACTCGCTGCCACTGTGGAGCTGAGATCACTGGACATGTCAAAAACGGGACGATGA
ACGTGTGAGCGACGGTGACACCTCGACTCTAGTGACCTGTACAGTTTTTGCCCTGCTACT

IleValGlyProArgThrCysArgAsnMetTrpSerGlyThrPheProIleAsnAlaTyr
4261 GGATCGTCGGTCCTAGGACCTGCAGGAACATGTGGAGTGGGACCTTCCCCATTAATGCCT
CCTAGCAGCCAGGATCCTGGACGTCCTTGTACACCTCACCCTGGAAGGGGTAATTACGGA

ThrThrGlyProCysThrProLeuProAlaProAsnTyrThrPheAlaLeuTrpArgVal
4321 ACACCACGGGCCCCTGTACCCCCCTTCCTGCGCCGAACTACACGTTCGCGCTATGGAGGG
TGTGGTGCCCGGGGACATGGGGGGAAGGACGCGGCTTGATGTGCAAGCGCGATACCTCCC

SerAlaGluGluTyrValGluIleArgGlnValGlyAspPheHisTyrValThrGlyMet
4381 TGTCTGCAGAGGAATATGTGGAGATAAGGCAGGTGGGGGACTTCCACTACGTGACGGGTA
ACAGACGTCTCCTTATACACCTCTATTCCGTCCACCCCCTGAAGGTGATGCACTGCCCAT

ThrThrAspAsnLeuLysCysProCysGlnValProSerProGluPhePheThrGluLeu
4441 TGACTACTGACAATCTCAAATGCCCGTGCCAGGTCCCATCGCCCGAATTTTTCACAGAAT
ACTGATGACTGTTAGAGTTTACGGGCACGGTCCAGGGTAGCGGGCTTAAAAAGTGTCTTA

AspGlyValArgLeuHisArgPheAlaProProCysLysProLeuLeuArgGluGluVal
4501 TGGACGGGGTGCGCCTACATAGGTTTGCGCCCCCCTGCAAGCCCTTGCTGCGGGAGGAGG
ACCTGCCCCACGCGGATGTATCCAAACGCGGGGGGACGTTCGGGAACGACGCCCTCCTCC

SerPheArgValGlyLeuHisGluTyrProValGlySerGlnLeuProCysGluProGlu
4561 TATCATTCAGAGTAGGACTCCACGAATACCCGGTAGGGTCGCAATTACCTTGCGAGCCCG
ATAGTAAGTCTCATCCTGAGGTGCTTATGGGCCATCCCAGCGTTAATGGAACGCTCGGGC

ProAspValAlaValLeuThrSerMetLeuThrAspProSerHisIleThrAlaGluAla
4621 AACCGGACGTGGCCGTGTTGACGTCCATGCTCACTGATCCCTCCCATATAACAGCAGAGG
TTGGCCTGCACCGGCACAACTGCAGGTACGAGTGACTAGGGAGGGTATATTGTCGTCTCC

AlaGlyArgArgLeuAlaArgGlySerProProSerValAlaSerSerSerAlaSerGln
4681 CGGCCGGGCGAAGGTTGGCGAGGGGATCACCCCCCTCTGTGGCCAGCTCCTCGGCTAGCC
GCCGGCCCGCTTCCAACCGCTCCCCTAGTGGGGGGAGACACCGGTCGAGGAGCCGATCGG

LeuSerAlaProSerLeuLysAlaThrCysThrAlaAsnHisAspSerProAspAlaGlu
4741 AGCTATCCGCTCCATCTCTCAAGGCAACTTGCACCGCTAACCATGACTCCCCTGATGCTG
TCGATAGGCGAGGTAGAGAGTTCCGTTGAACGTGGCGATTGGTACTGAGGGGACTACGAC

LeuIleGluAlaAsnLeuLeuTrpArgGlnGluMetGlyGlyAsnIleThrArgValGlu
4801 AGCTCATAGAGGCCAACCTCCTATGGAGGCAGGAGATGGGCGGCAACATCACCAGGGTTG
TCGAGTATCTCCGGTTGGAGGATACCTCCGTCCTCTACCCGCCGTTGTAGTGGTCCCAAC

SerGluAsnLysValValIleLeuAspSerPheAspProLeuValAlaGluGluAspGlu
4861 AGTCAGAAAACAAAGTGGTGATTCTGGACTCCTTCGATCCGCTTGTGGCGGAGGAGGACG
TCAGTCTTTTGTTTCACCACTAAGACCTGAGGAAGCTAGGCGAACACCGCCTCCTCCTGC

FIG. 32-5

```
      ArgGluIleSerValProAlaGluIleLeuArgLysSerArgArgPheAlaGlnAlaLeu
4921  AGCGGGAGATCTCCGTACCCGCAGAAATCCTGCGGAAGTCTCGGAGATTCGCCCAGGCCC
      TCGCCCTCTAGAGGCATGGGCGTCTTTAGGACGCCTTCAGAGCCTCTAAGCGGGTCCGGG

      ProValTrpAlaArgProAspTyrAsnProProLeuValGluThrTrpLysLysProAsp
4981  TGCCCGTTTGGGCGCGGCCGGACTATAACCCCCCGCTAGTGGAGACGTGGAAAAAGCCCG
      ACGGGCAAACCCGCGCCGGCCTGATATTGGGGGGCGATCACCTCTGCACCTTTTTCGGGC

      TyrGluProProValValHisGlyCysProLeuProProProLysSerProProValPro
5041  ACTACGAACCACCTGTGGTCCATGGCTGTCCGCTTCCACCTCCAAAGTCCCCTCCTGTGC
      TGATGCTTGGTGGACACCAGGTACCGACAGGCGAAGGTGGAGGTTTCAGGGGAGGACACG

      ProProArgLysLysArgThrValValLeuThrGluSerThrLeuSerThrAlaLeuAla
5101  CTCCGCCTCGGAAGAAGCGGACGGTGGTCCTCACTGAATCAACCCTATCTACTGCCTTGG
      GAGGCGGAGCCTTCTTCGCCTGCCACCAGGAGTGACTTAGTTGGGATAGATGACGGAACC

      GluLeuAlaThrArgSerPheGlySerSerSerThrSerGlyIleThrGlyAspAsnThr
5161  CCGAGCTCGCCACCAGAAGCTTTGGCAGCTCCTCAACTTCCGGCATTACGGGCGACAATA
      GGCTCGAGCGGTGGTCTTCGAAACCGTCGAGGAGTTGAAGGCCGTAATGCCCGCTGTTAT

      ThrThrSerSerGluProAlaProSerGlyCysProProAspSerAspAlaGluSerTyr
5221  CGACAACATCCTCTGAGCCCGCCCCTTCTGGCTGCCCCCCCGACTCCGACGCTGAGTCCT
      GCTGTTGTAGGAGACTCGGGCGGGGAAGACCGACGGGGGGGCTGAGGCTGCGACTCAGGA

      SerSerMetProProLeuGluGlyGluProGlyAspProAspLeuSerAspGlySerTrp
5281  ATTCCTCCATGCCCCCCCTGGAGGGGGAGCCTGGGGATCCGGATCTTAGCGACGGGTCAT
      TAAGGAGGTACGGGGGGGACCTCCCCCTCGGACCCCTAGGCCTAGAATCGCTGCCCAGTA

      SerThrValSerSerGluAlaAsnAlaGluAspValValCysCysSerMetSerTyrSer
5341  GGTCAACGGTCAGTAGTGAGGCCAACGCGGAGGATGTCGTGTGCTGCTCAATGTCTTACT
      CCAGTTGCCAGTCATCACTCCGGTTGCGCCTCCTACAGCACACGACGAGTTACAGAATGA

      TrpThrGlyAlaLeuValThrProCysAlaAlaGluGluGlnLysLeuProIleAsnAla
5401  CTTGGACAGGCGCACTCGTCACCCCGTGCGCCGCGGAAGAACAGAAACTGCCCATCAATG
      GAACCTGTCCGCGTGAGCAGTGGGGCACGCGGCGCCTTCTTGTCTTTGACGGGTAGTTAC

      LeuSerAsnSerLeuLeuArgHisHisAsnLeuValTyrSerThrThrSerArgSerAla
5461  CACTAAGCAACTCGTTGCTACGTCACCACAATTTGGTGTATTCCACCACCTCACGCAGTG
      GTGATTCGTTGAGCAACGATGCAGTGGTGTTAAACCACATAAGGTGGTGGAGTGCGTCAC

      CysGlnArgGlnLysLysValThrPheAspArgLeuGlnValLeuAspSerHisTyrGln
5521  CTTGCCAAAGGCAGAAGAAAGTCACATTTGACAGACTGCAAGTTCTGGACAGCCATTACC
      GAACGGTTTCCGTCTTCTTTCAGTGTAAACTGTCTGACGTTCAAGACCTGTCGGTAATGG

      AspValLeuLysGluValLysAlaAlaAlaSerLysValLysAlaAsnLeuLeuSerVal
5581  AGGACGTACTCAAGGAGGTTAAAGCAGCGGCGTCAAAAGTGAAGGCTAACTTGCTATCCG
      TCCTGCATGAGTTCCTCCAATTTCGTCGCCGCAGTTTTCACTTCCGATTGAACGATAGGC

      GluGluAlaCysSerLeuThrProProHisSerAlaLysSerLysPheGlyTyrGlyAla
5641  TAGAGGAAGCTTGCAGCCTGACGCCCCCACACTCAGCCAAATCCAAGTTTGGTTATGGGG
      ATCTCCTTCGAACGTCGGACTGCGGGGGTGTGAGTCGGTTTAGGTTCAAACCAATACCCC

      LysAspValArgCysHisAlaArgLysAlaValThrHisIleAsnSerValTrpLysAsp
5701  CAAAAGACGTCCGTTGCCATGCCAGAAAGGCCGTAACCCACATCAACTCCGTGTGGAAAG
      GTTTTCTGCAGGCAACGGTACGGTCTTTCCGGCATTGGGTGTAGTTGAGGCACACCTTTC

      LeuLeuGluAspAsnValThrProIleAspThrThrIleMetAlaLysAsnGluValPhe
5761  ACCTTCTGGAAGACAATGTAACACCAATAGACACTACCATCATGGCTAAGAACGAGGTTT
      TGGAAGACCTTCTGTTACATTGTGGTTATCTGTGATGGTAGTACCGATTCTTGCTCCAAA

      CysValGlnProGluLysGlyGlyArgLysProAlaArgLeuIleValPheProAspLeu
5821  TCTGCGTTCAGCCTGAGAAGGGGGGTCGTAAGCCAGCTCGTCTCATCGTGTTCCCCGATC
      AGACGCAAGTCGGACTCTTCCCCCCAGCATTCGGTCGAGCAGAGTAGCACAAGGGGCTAG

      GlyValArgValCysGluLysMetAlaLeuTyrAspValValThrLysLeuProLeuAla
```

FIG. 32-6

```
5881  TGGGCGTGCGCGTGTGCGAAAAGATGGCTTTGTACGACGTGGTTACAAAGCTCCCCTTGG
      ACCCGCACGCGCACACGCTTTTCTACCGAAACATGCTGCACCAATGTTTCGAGGGGAACC

       ValMetGlySerSerTyrGlyPheGlnTyrSerProGlyGlnArgValGluPheLeuVal
5941  CCGTGATGGGAAGCTCCTACGGATTCCAATACTCACCAGGACAGCGGGTTGAATTCCTCG
      GGCACTACCCTTCGAGGATGCCTAAGGTTATGAGTGGTCCTGTCGCCCAACTTAAGGAGC

       GlnAlaTrpLysSerLysLysThrProMetGlyPheSerTyrAspThrArgCysPheAsp
6001  TGCAAGCGTGGAAGTCCAAGAAAACCCCAATGGGGTTCTCGTATGATACCCGCTGCTTTG
      ACGTTCGCACCTTCAGGTTCTTTTGGGGTTACCCCAAGAGCATACTATGGGCGACGAAAC

       SerThrValThrGluSerAspIleArgThrGluGluAlaIleTyrGlnCysCysAspLeu
6061  ACTCCACAGTCACTGAGAGCGACATCCGTACGGAGGAGGCAATCTACCAATGTTGTGACC
      TGAGGTGTCAGTGACTCTCGCTGTAGGCATGCCTCCTCCGTTAGATGGTTACAACACTGG

       AspProGlnAlaArgValAlaIleLysSerLeuThrGluArgLeuTyrValGlyGlyPro
6121  TCGACCCCCAAGCCCGCGTGGCCATCAAGTCCCTCACCGAGAGGCTTTATGTTGGGGGCC
      AGCTGGGGGTTCGGGCGCACCGGTAGTTCAGGGAGTGGCTCTCCGAAATACAACCCCCGG

       LeuThrAsnSerArgGlyGluAsnCysGlyTyrArgArgCysArgAlaSerGlyValLeu
6181  CTCTTACCAATTCAAGGGGGGAGAACTGCGGCTATCGCAGGTGCCGCGCGAGCGGCGTAC
      GAGAATGGTTAAGTTCCCCCCTCTTGACGCCGATAGCGTCCACGGCGCGCTCGCCGCATG

       ThrThrSerCysGlyAsnThrLeuThrCysTyrIleLysAlaArgAlaAlaCysArgAla
6241  TGACAACTAGCTGTGGTAACACCCTCACTTGCTACATCAAGGCCCGGGCAGCCTGTCGAG
      ACTGTTGATCGACACCATTGTGGGAGTGAACGATGTAGTTCCGGGCCCGTCGGACAGCTC

       AlaGlyLeuGlnAspCysThrMetLeuValCysGlyAspAspLeuValValIleCysGlu
6301  CCGCAGGGCTCCAGGACTGCACCATGCTCGTGTGTGGCGACGACTTAGTCGTTATCTGTG
      GGCGTCCCGAGGTCCTGACGTGGTACGAGCACACACCGCTGCTGAATCAGCAATAGACAC

       SerAlaGlyValGlnGluAspAlaAlaSerLeuArgAlaPheThrGluAlaMetThrArg
6361  AAAGCGCGGGGGTCCAGGAGGACGCGGCGAGCCTGAGAGCCTTCACGGAGGCTATGACCA
      TTTCGCGCCCCCAGGTCCTCCTGCGCCGCTCGGACTCTCGGAAGTGCCTCCGATACTGGT

       TyrSerAlaProProGlyAspProProGlnProGluTyrAspLeuGluLeuIleThrSer
6421  GGTACTCCGCCCCCCCTGGGGACCCCCCACAACCAGAATACGACTTGGAGCTCATAACAT
      CCATGAGGCGGGGGGGACCCCTGGGGGGTGTTGGTCTTATGCTGAACCTCGAGTATTGTA

       CysSerSerAsnValSerValAlaHisAspGlyAlaGlyLysArgValTyrTyrLeuThr
6481  CATGCTCCTCCAACGTGTCAGTCGCCCACGACGGCGCTGGAAAGAGGGTCTACTACCTCA
      GTACGAGGAGGTTGCACAGTCAGCGGGTGCTGCCGCGACCTTTCTCCCAGATGATGGAGT

       ArgAspProThrThrProLeuAlaArgAlaAlaTrpGluThrAlaArgHisThrProVal
6541  CCCGTGACCCTACAACCCCCCTCGCGAGAGCTGCGTGGGAGACAGCAAGACACACTCCAG
      GGGCACTGGGATGTTGGGGGGAGCGCTCTCGACGCACCCTCTGTCGTTCTGTGTGAGGTC

       AsnSerTrpLeuGlyAsnIleIleMetPheAlaProThrLeuTrpAlaArgMetIleLeu
6601  TCAATTCCTGGCTAGGCAACATAATCATGTTTGCCCCCACACTGTGGGCGAGGATGATAC
      AGTTAAGGACCGATCCGTTGTATTAGTACAAACGGGGGTGTGACACCCGCTCCTACTATG

       MetThrHisPhePheSerValLeuIleAlaArgAspGlnLeuGluGlnAlaLeuAspCys
6661  TGATGACCCATTTCTTTAGCGTCCTTATAGCCAGGGACCAGCTTGAACAGGCCCTCGATT
      ACTACTGGGTAAAGAAATCGCAGGAATATCGGTCCCTGGTCGAACTTGTCCGGGAGCTAA

       GluIleTyrGlyAlaCysTyrSerIleGluProLeuAspLeuProProIleIleGlnArg
6721  GCGAGATCTACGGGGCCTGCTACTCCATAGAACCACTTGATCTACCTCCAATCATTCAAA
      CGCTCTAGATGCCCCGGACGATGAGGTATCTTGGTGAACTAGATGGAGGTTAGTAAGTTT

       Leu
6781  GACTC
      CTGAG
```

FIG. 32-7

FIG. 33 LEGEND

Lane Number	Chimp Reference Number	Infection Type	Sample date (days) (0=inoculation day)	ALT (alanine) aminotransferase level in sera) mu/ml)
1	1	NANB	0	9
2	1	NANB	76	71
3	1	NANB	118	19
4	1	NANB	154	N/A
5	2	NANB	0	5
6	2	NANB	21	52
7	2	NANB	73	13
8	2	NANB	138	N/A
9	3	NANB	0	8
10	3	NANB	43	205
11	3	NANB	53	14
12	3	NANB	159	6
13	4	NANB	-3	11
14	4	NANB	55	132
15	4	NANB	83	N/A
16	4	NANB	140	N/A
17	5	HAV	0	4
18	5	HAV	25	147
19	5	HAV	40	18
20	5	HAV	268	5
21	6	HAV	-8	N/A
22	6	HAV	15	106
23	6	HAV	41	10
24	6	HAV	129	N/A
26	7	HAV	0	7
27	7	HAV	22	83
28	7	HAV	115	5
29	7	HAV	139	N/A
30	8	HAV	0	15
31	8	HAV	26	130
32	8	HAV	74	8
33	8	HAV	205	5
34	9	HBV	-290	N/A
35	9	HBV	379	9
36	9	HBV	435	6
37	10	HBV	0	8
38	10	HBV	111-118 (pool)	96-156 (pool)
39	10	HBV	205	9
40	10	HBV	240	13
41	11	HBV	0	11
42	11	HBV	28-56 (pool)	8-100 (pool)
43	11	HBV	169	9
44	11	HBV	223	10

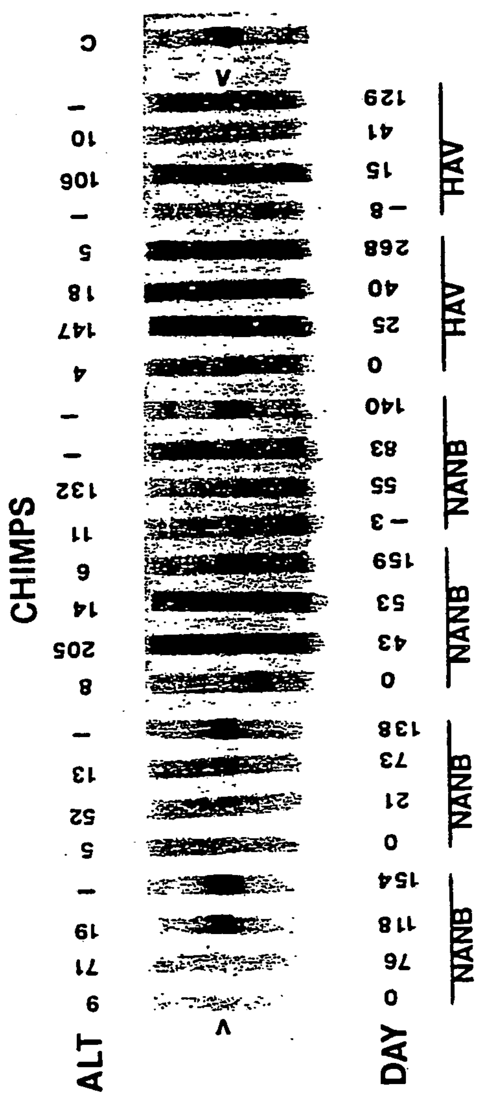

FIG. 33-1

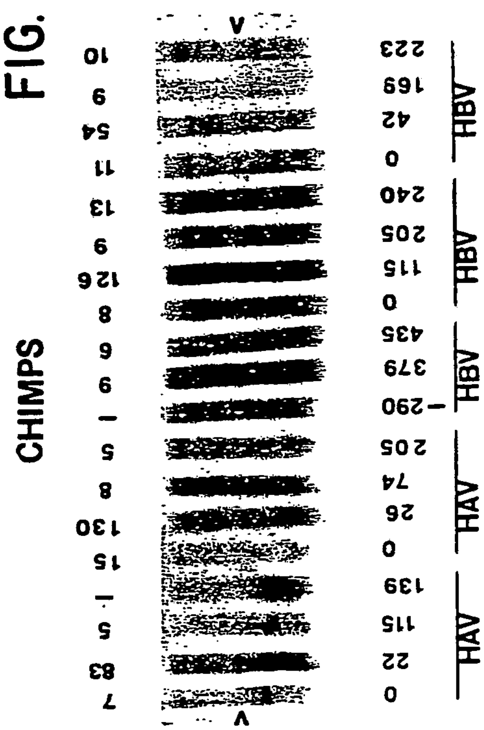

FIG. 33-2

FIG. 34 LEGEND

Lane Number	Patient Reference Number	Diagnosis	ALT Level (mu/ml)
1	1[1]	NANB	1354
2	1[1]	NANB	31
3	2[1]	NANB	14
4	2[1]	NANB	79
5	2[1]	NANB	26
6	3[1]	NANB	78
7	3[1]	NANB	87
8	3[1]	NANB	25
9	4[1]	NANB	60
10	4[1]	NANB	13
11	5[1]	NANB	298
12	5[1]	NANB	101
13	6[1]	NANB	474
14	6[1]	NANB	318
15	7[1]	NANB	20
16	7[1]	NANB	163
17	8[1]	NANB	44
18	8[1]	NANB	50
19	9	NANB	N/A
20	10	NANB	N/A
21	11	NANB	N/A
22	12	Normal	N/A
23	13	Normal	N/A
24	14	Normal	N/A
26	30174	Normal	N/A
27	30105	Normal	N/A
28	30072	Normal	N/A
29	30026	Normal	N/A
30	30146	Normal	N/A
31	30250	Normal	N/A
32	30071	Normal	N/A
33	15	AcuteHAV	N/A
34	16	AcuteHAV	N/A
35	17	AcuteHAV	N/A
36	18	AcuteHAV	N/A
37	48088	AcuteHAV	N/A
38	47288	AcuteHAV	N/A
39	47050	AcuteHAV	N/A
40	46997	AcuteHAV	N/A
41	19	Convalescent HBV	N/A
42	20	(anti-HBSag+ve;	N/A
43	21	anti-HBCag+ve)	N/A
44	22	(anti-HBSag+ve;	N/A
45	23	anti-HBCag+ve)	N/A
46	24	(anti-HBSag+ve;	N/A
47	25	anti-HBCag+ve)	N/A
48	26	(anti-HBSag+ve;	N/A
49	27	anti-HBSag+ve)	N/A

[1]Sequential serum samples were assayed from these patients

FIG. 34-1

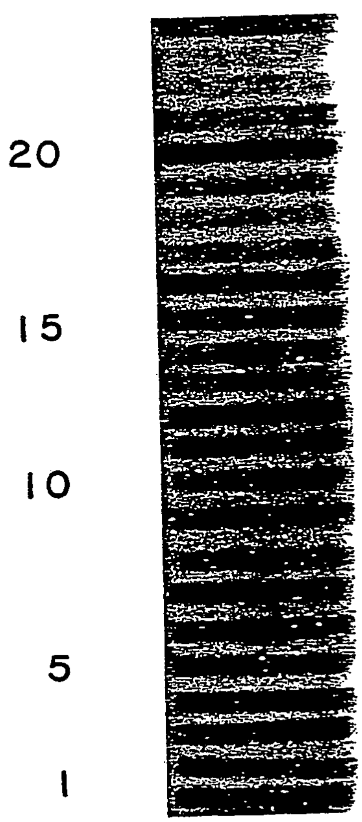

FIG. 34-2

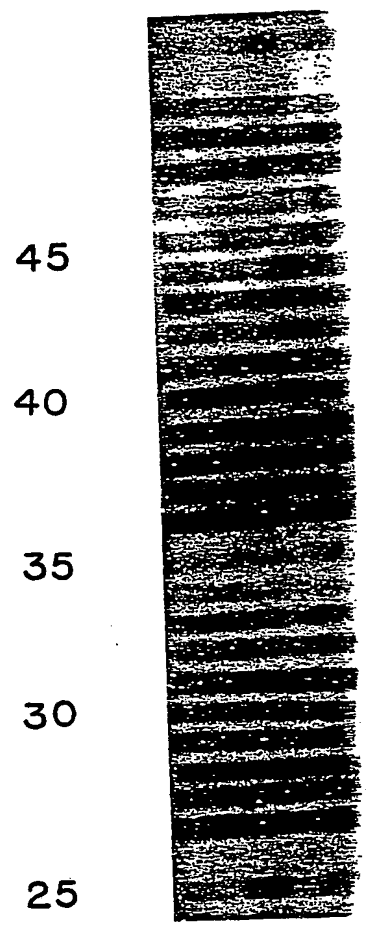

FIG. 35

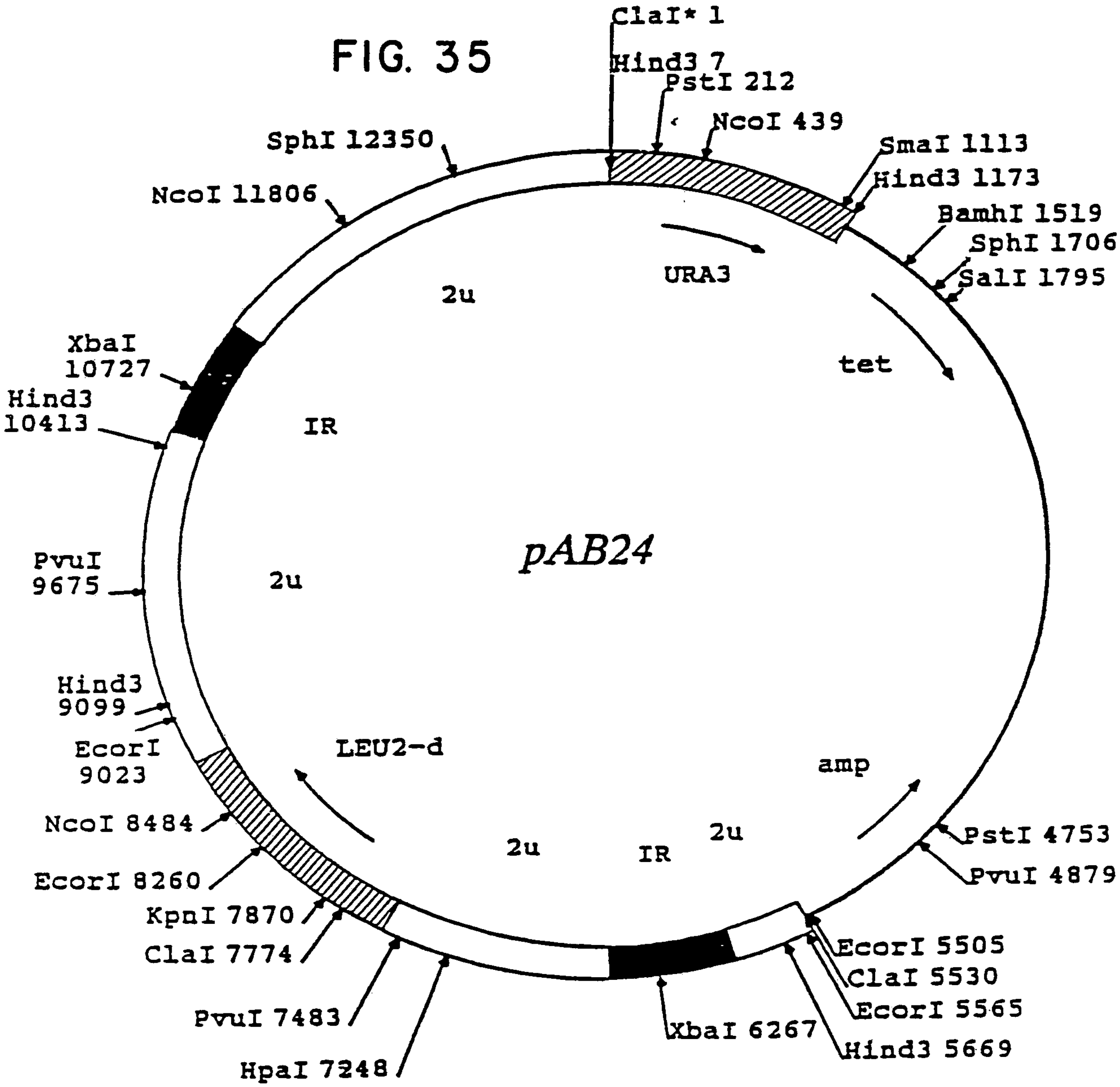
ClaI* 1
Hind3 7
PstI 212
NcoI 439
SmaI 1113
Hind3 1173
BamhI 1519
SphI 1706
SalI 1795
SphI 12350
NcoI 11806
2u
URA3
tet
XbaI
10727
Hind3
10413
IR
pAB24
PvuI
9675
2u
Hind3
9099
EcorI
9023
LEU2-d
amp
NcoI 8484
2u
IR
2u
PstI 4753
PvuI 4879
EcorI 8260
KpnI 7870
ClaI 7774
EcorI 5505
ClaI 5530
EcorI 5565
PvuI 7483
XbaI 6267
Hind3 5669
HpaI 7248

FIG. 36-1 COOH-terminus of SOD-C100 Fusion Polypeptide

```
     --------SOD-----------COOH][--adaptor----][NANBHpolypeptide>
     AlaCysGlyValIleGlyIleAlaGlnAsnLeuGlyIleArgAspAlaHisPheLeuSer
  1  GCTTGTGGTGTAATTGGGATCGCCCAGAATTTGGGAATTCGGGATGCCCACTTTCTATCC
     CGAACACCACATTAACCCTAGCGGGTCTTAAACCCTTAAGCCCTACGGGTGAAAGATAGG

     >>>>>>>>>>>>>>>>>>>>>>
     GlnThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGlnAlaThrValCys
 61  CAGACAAAGCAGAGTGGGGAGAACCTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGC
     GTCTGTTTCGTCTCACCCCTCTTGGAAGGAATGGACCATCGCATGGTTCGGTGGCACACG

     AlaArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLysCysLeuIleArgLeu
121  GCTAGGGCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATTCGCCTC
     CGATCCCGAGTTCGGGGAGGGGGTAGCACCCTGGTCTACACCTTCACAAACTAAGCGGAG

     LysProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAlaValGlnAsnGlu
181  AAGCCCACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAA
     TTCGGGTGGGAGGTACCCGGTTGTGGGGACGATATGTCTGACCCGCGACAAGTCTTACTT

     IleThrLeuThrHisProValThrLysTyrIleMetThrCysMetSerAlaAspLeuGlu
241  ATCACCCTGACGCACCCAGTCACCAAATACATCATGACATGCATGTCGGCCGACCTGGAG
     TAGTGGGACTGCGTGGGTCAGTGGTTTATGTAGTACTGTACGTACAGCCGGCTGGACCTC

     ValValThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeuAlaAlaTyrCys
301  GTCGTCACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTTTGGCCGCGTATTGC
     CAGCAGTGCTCGTGGACCCACGAGCAACCGCCGCAGGACCGACGAAACCGGCGCATAACG

     LeuSerThrGlyCysValValIleValGlyArgValValLeuSerGlyLysProAlaIle
361  CTGTCAACAGGCTGCGTGGTCATAGTGGGCAGGGTCGTCTTGTCCGGGAAGCCGGCAATC
     GACAGTTGTCCGACGCACCAGTATCACCCGTCCCAGCAGAACAGGCCCTTCGGCCGTTAG

     IleProAspArgGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHis
421  ATACCTGACAGGGAAGTCCTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCAC
     TATGGACTGTCCCTTCAGGAGATGGCTCTCAAGCTACTCTACCTTCTCACGAGAGTCGTG

     LeuProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGly
481  TTACCGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAAGGCCCTCGGC
     AATGGCATGTAGCTCGTTCCCTACTACGAGCGGCTCGTCAAGTTCGTCTTCCGGGAGCCG

     LeuLeuGlnThrAlaSerArgGlnAlaGluValIleAlaProAlaValGlnThrAsnTrp
541  CTCCTGCAGACCGCGTCCCGTCAGGCAGAGGTTATCGCCCCTGCTGTCCAGACCAACTGG
     GAGGACGTCTGGCGCAGGGCAGTCCGTCTCCAATAGCGGGGACGACAGGTCTGGTTGACC

     GlnLysLeuGluThrPheTrpAlaLysHisMetTrpAsnPheIleSerGlyIleGlnTyr
601  CAAAAACTCGAGACCTTCTGGGCGAAGCATATGTGGAACTTCATCAGTGGGATACAATAC
     GTTTTTGAGCTCTGGAAGACCCGCTTCGTATACACCTTGAAGTAGTCACCCTATGTTATG

     LeuAlaGlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeuMetAlaPheThr
661  TTGGCGGGCTTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACA
     AACCGCCCGAACAGTTGCGACGGACCATTGGGGCGGTAACGAAGTAACTACCGAAAATGT

     AlaAlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsnIleLeuGlyGly
721  GCTGCTGTCACCAGCCCACTAACCACTAGCCAAACCCTCCTCTTCAACATATTGGGGGGG
     CGACGACAGTGGTCGGGTGATTGGTGATCGGTTTGGGAGGAGAAGTTGTATAACCCCCCC

     TrpValAlaAlaGlnLeuAlaAlaProGlyAlaAlaThrAlaPheValGlyAlaGlyLeu
781  TGGGTGGCTGCCCAGCTCGCCGCCCCCGGTGCCGCTACTGCCTTTGTGGGCGCTGGCTTA
     ACCCACCGACGGGTCGAGCGGCGGGGGCCACGGCGATGACGGAAACACCCGCGACCGAAT

     AlaGlyAlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAspIleLeuAlaGly
841  GCTGGCGCCGCCATCGGCAGTGTTGGACTGGGGAAGGTCCTCATAGACATCCTTGCAGGG
```

```
      CGACCGCGGCGGTAGCCGTCACAACCTGACCCCTTCCAGGAGTATCTGTAGGAACGTCCC

      TyrGlyAlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSerGlyGluValPro
 901  TATGGCGCGGGCGTGGCGGGAGCTCTTGTGGCATTCAAGATCATGAGCGGTGAGGTCCCC
      ATACCGCGCCCGCACCGCCCTCGAGAACACCGTAAGTTCTAGTACTCGCCACTCCAGGGG

      SerThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGlyAlaLeuValVal
 961  TCCACGGAGGACCTGGTCAATCTACTGCCCGCCATCCTCTCGCCCGGAGCCCTCGTAGTC
      AGGTGCCTCCTGGACCAGTTAGATGACGGGCGGTAGGAGAGCGGGCCTCGGGAGCATCAG

      GlyValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGluGlyAlaValGln
1021  GGCGTGGTCTGTGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAGGGGGCAGTGCAG
      CCGCACCAGACACGTCGTTATGACGCGGCCGTGCAACCGGGCCCGCTCCCCCGTCACGTC

                               <<<<<<<<<<<<<<<<<<<<NANBH][---extra
      TrpMetAsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSerProValHisHis
1081  TGGATGAACCGGCTGATAGCCTTCGCCTCCCGGGGGAACCATGTTTCCCCAGTCCATCAT
      ACCTACTTGGCCGACTATCGGAAGCGGAGGGCCCCCTTGGTACAAAGGGGTCAGGTAGTA

      ------]
      LysArgOP
1141  AAGCGTTGACGCTCCCTACGGGTGGACTGTGGAGAGACAGGGCACTGCTAAGGCCCAAAT
      TTCGCAACTGCGAGGGATGCCCACCTGACACCTCTCTGTCCCGTGACGATTCCGGGTTTA

1201  CTCAGCCATGCATCGAGGGGTACAATCCGTATGGCCAACAACTAGCGCGTACGTAAAGTC
      GAGTCGGTACGTAGCTCCCCATGTTAGGCATACCGGTTGTTGATCGCGCATGCATTTCAG

1261  TCCTTTCTCGATGGTCCATACCTTAGATGCGTTAGCATTAATCCGAATTC
      AGGAAAGAGCTACCAGGTATGGAATCTACGCAATCGTAATTAGGCTTAAG
```

FIG. 36-2

FIG. 37a

FIG. 37b

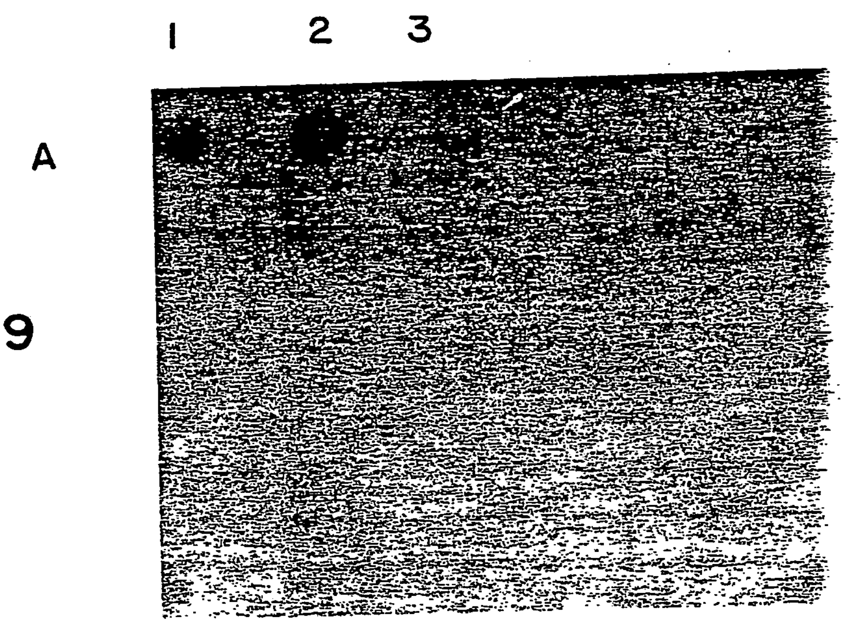

FIG. 39

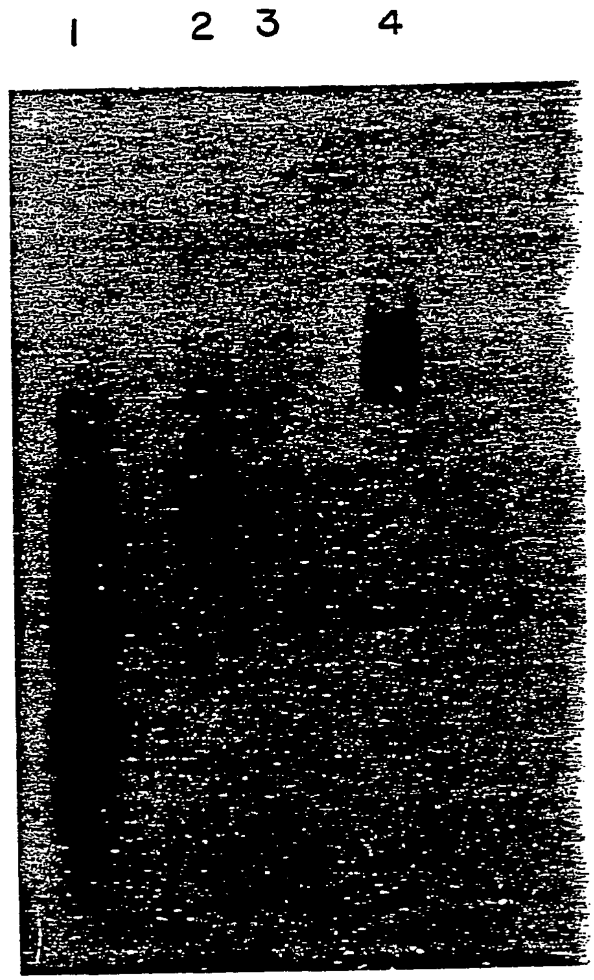

FIG. 38

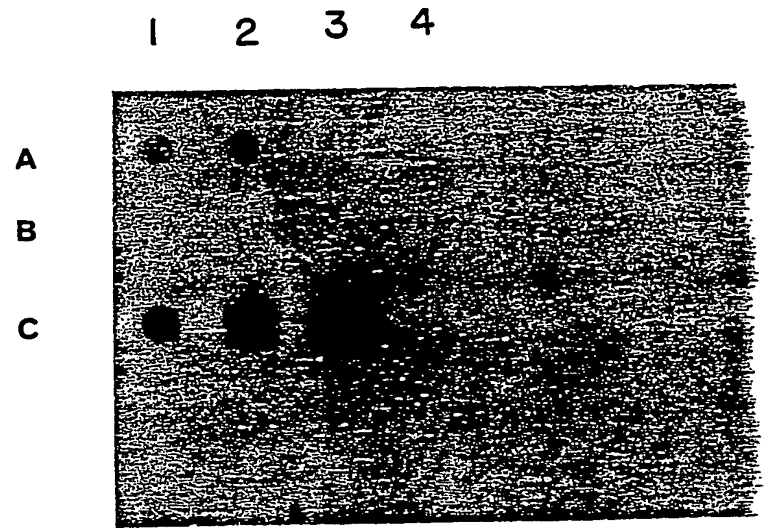

FIG. 40

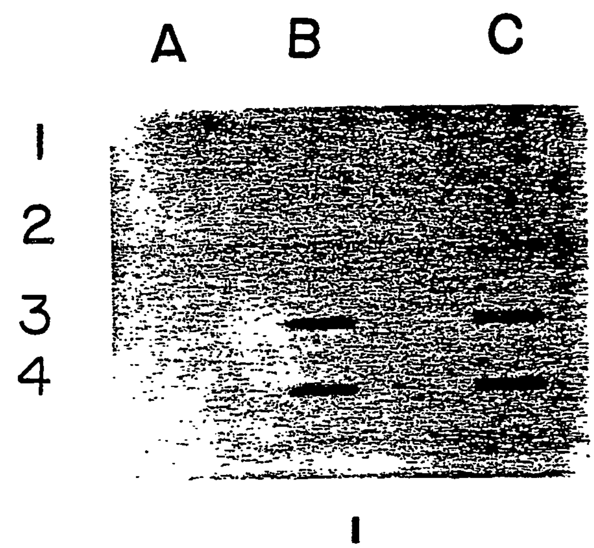

FIG. 41a

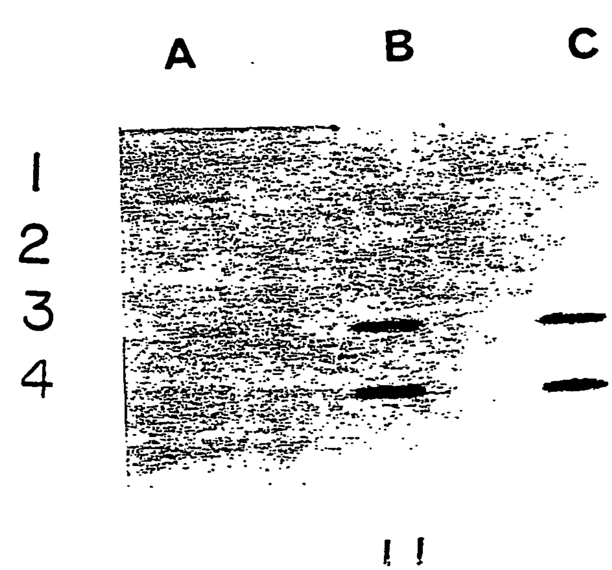

FIG. 41b

FIG. 41-1

Homology between the HCV polypeptide encoded by combined ORF of clones 14i through 39c) and the non-structural protein of the Dengue flavivirus(MNWVD1).

```
                   10        20        30        40        50
HCV       EYVVLLFLLLADARVCSCLWMMLLISQAEAALENLVILNAASLAGTHGLVSFLVFFCFA

MNWVD1   AVSFVTLITGNMSFRDLGRVMVMVGATMTDDIGMGVTYLALLAAFKVRPTFAAGLLLRKL
               130       140       150       160       170       180

         60        70        80        90       100       110
HCV      WYLKGKWVPGAVYTFYGMWPLLLLLLALPQRAYALDTEVAASCGGVVLVGLMALTLSPYY

MNWVD1   TSKELMMTTIGIVLLSQSTIPETILELTDALALGMMVLKMVRKMEKYQLAVTIMAILCVP
               190       200       210       220       230       240

        120       130       140       150       160       170
HCV      KRYISWCLWWLQYFLTRVEAQLHVWIPPLNVRGGRDAVILLMCAVHPTLVFDITKLLLAV
                :  .  .  .  :  .   ......  .  ....  ..  :  .  ...::  .:  .:.:   .
MNWVD1   NAVILQNAWKVSCTILAVVSVSPLFLTSSQQKADWIPLALTIKGLNPTAIF-LTTLSRTN
               250       260       270       280       290

        180       190       200        210       220       230
HCV      FGPLWILQASLLKVPYF-VRVQGLLRF-CALARKMIGGHYVQMVIIKLGALTGTYVYNHL
          .  :  :.....  :     .  .  .::.   ...   ...:   .   .    .  :..  ...     ..
MNWVD1   KKRSWPLNEAIMAVGMVSILASSLLKNDIPMTGPLVAGGLLTVCYV-LTGRSADLELERA
        300       310       320       330       340       350

          240       250       260       270       280       290
HCV      TPLRDWAHNGLRDLAVAVEPVVFSQMETKLITWGADTAACGDIINGLPVSARRGREILLG
          .  ..  :....      .  .  ....:.  ....  .  ......      .:.      .   .  .:     :.
MNWVD1   ADVK-WEDQAEISGSSPILSITISE-DGSMSIKNEEEEQTLTILIRTGLLVISG---LFP
          360          370       380       390       400       410

          300       310       320       330       340       350
HCV      PADGMVSKGWRLLAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATC
          .   ...  .:  :  .   ..  :    ...     ..  ..  ....     .   .   .   .  .:.   ...
MNWVD1   VSIPITAAAWYLWEVKKQRAGVLWDVPSPPPVGKAELEDGAYRIKQKGILGYSQIGAGVY
             420       430       440       450       460       470

          360       370       380       390           400       410
HCV      INGVCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLV----GWPAPQGSRSLTPCTCGSSD
          .:.   :..:  .  .  ..      ..  .  ...:..:::     ::      .  ..  .        .  .
MNWVD1   KEGTFHTMWHVTRGAVLMHKGKRIEPSWADVKKDLVSCGGGWKLEGEWKEGEEVQVLALE
             480       490       500       510       520       530

             420       430       440       450       460       470
HCV      LYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGV
          .:  ...  :      ...   ..  ...  ....     :.::.:..      :..::..  ..:  ::.
MNWVD1   PGKNPRAVQTKPGLFKTN--AGTIGAVSLDFSPGTSGSPIIDKKGKVVGLYGNGVVTRSG
             540       550       560       570       580       590

             480       490       500       510       520         530
HCV      AKAVDFIPVENLETTMRSPVFTDNSSPPVVPQSFQVAHLHAPTGSGKS--TKVPAAYAAQ
         :  .  ..  ..:.     ..  .:  ..:.   .        ...   .   .::.  .:.:.            .
MNWVD1   AYVSAIAQTEK--SIEDNPEIEDDIFRK---RKLTIMDLHPGAGKTKRYLPAIVRGAIKR
               600       610          620       630       640

               540       550       560       570       580
```

```
HCV      GYKVLVLNPS--VAATLGFGAYMSKAHGIDPNIRTGVRTITTGSPITYSTYGKFLADGGC
         : ..:.:.:.  :::...  .       .  .:.::..  . ..  .  . ..:... :  .
MNWVD1   GLRTLILAPTRVVAAEMEEALRGLPIRYQTPAIRAEHTGREIVDLMCHATFTMRLL-SPV
          650       660       670       680       690       700

         590       600       610       620       630       640
HCV      SGGAYDIIICDECHSTDATSILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEV
         .  .X..:: :: : ::..:: . :  . ...:  :.   .. ::::::::  . : : ^ v.
MNWVD1   RVPNYNLIIMDEAHFTDPASIAARGYISTRVE-MGEAAGIFMTATPPGSRD-PFPQSNAP
           710       720       730       740       750       760

         650       660       670       680       690       700
HCV      ALSTTGEIPFYGKAIPLEVIKGGRHLIFCHSKKKCDELAAKLVALGINAVAYYRGLDVSV
          .... :::
MNWVD1   IMDEEREIPERSWSSGHEWVTDFKGKTVWFVPSIKAGNDTAACLRKNGKKVTQLSRKTFD
            770       780       790       800       810       820

         710       720       730       740       750       760
HCV      IPTSGDVVVVATDALMTGYTGDFDSVIDCNTCVTQTVDFSLDPTFTIETITLPQDAVSRT

MNWVD1   SEYVKTRTNDWNFVVTTDISEMGANFKAERVIDPRRCMKPVILTDGEERVILAGPMPVTH
            830       840       850       860       870       880

         770       780       790       800       810       820
HCV      QRRGRTGRGKPGIYRFVAPGERPSGMFDSSVLCECYDAGCAWYELTPAETTVRLRAYMNT

MNWVD1   SS
```

FIG. 41-2

FIG. 43

DISTRIBUTION OF RANDOM SAMPLES

C100-3 Ag ELISA Preclinical Kit

416ng C100/WELL, 2 HRS 37°C, 20ul SAMPLE

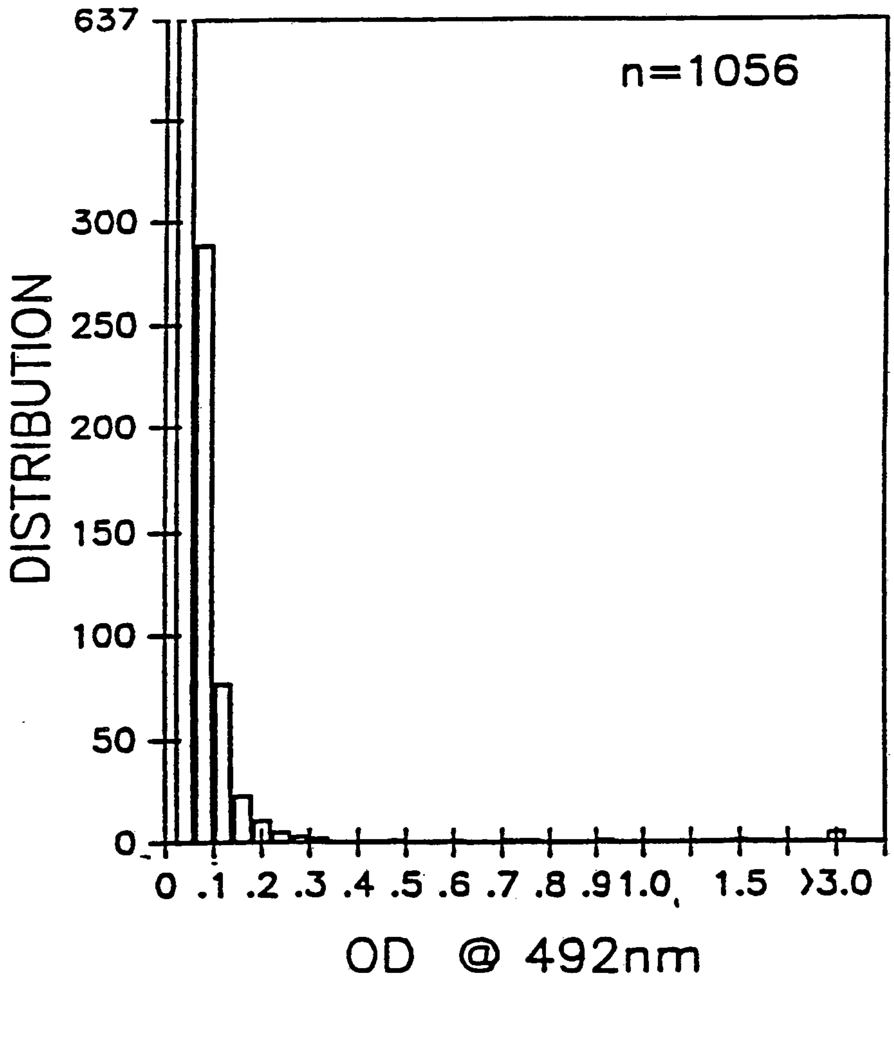

FIG. 44

Distribution of O.D. Values for Random Blood Donor Samples Tested with Two ELISA Configurations

C100–3 Ag ELISA MoAB vs Polyclonal

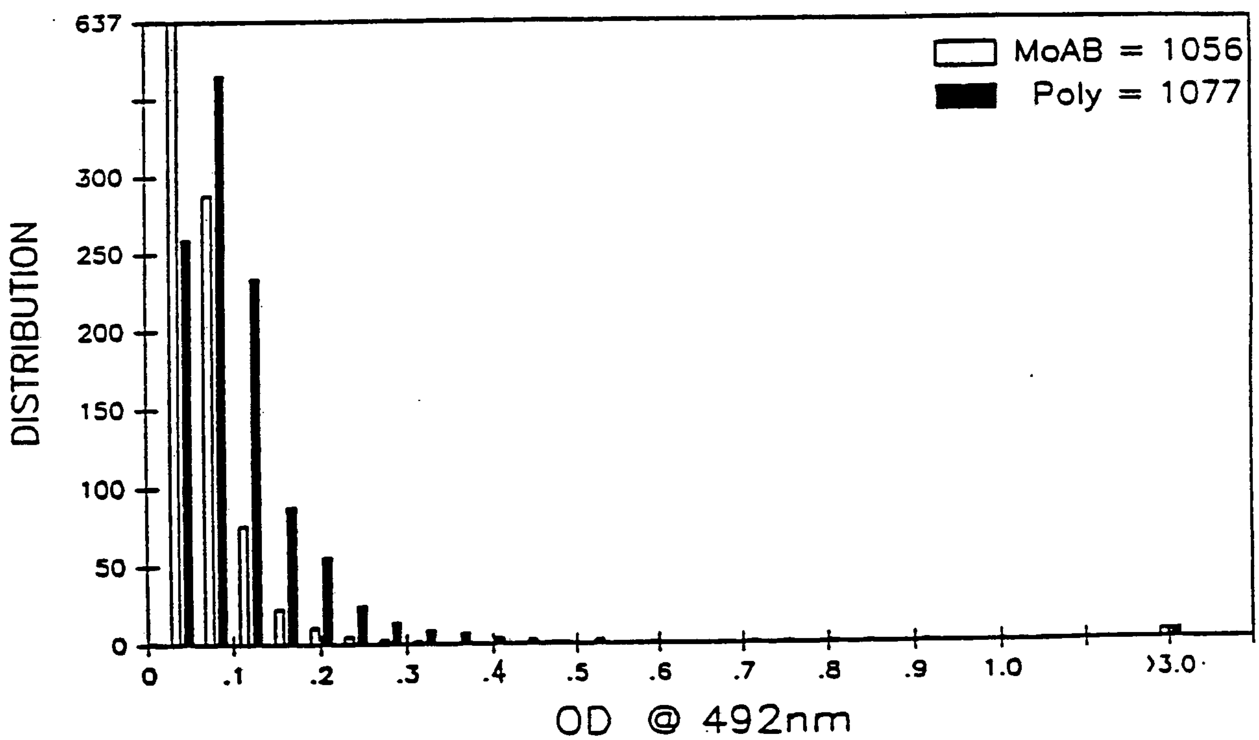

FIG. 45

Name	Common Sequence	Variable Sequence
5'-3-1	AAGCTTGATCGAATTC	CGATCTTGC
-2		CGATCCTGC
-3		CGATCATGC
-4		CGATCGTGC
-5		CGAAGTTGC
-6		CGAAGCTGC
-7		AGATCTTGC
-8		AGATCCTGC
-9		AGATCATGC
-10		AGATCGTGC
-11		AGAAGTTGC
-12		AGAAGCTGC
-13		CGATCTTGT
-14		CGATCCTGT
-15		CGATCATGT
-16		CGATCGTGT
-17		CGAAGTTGT
-18		CGAAGCTGT
-19		AGATCTTGT
-20		AGATCCTGT
-21		AGATCATGT
-22		AGATCGTGT
-23		AGAAGTTGT
-24		AGAAGCTGT
-25		CGCTCTTGC
-26		CGCTCCTGC
-27		CGCTCATGC
-28		CGCTCGTGC
-29		CGCAGTTGC
-30		CGCAGCTGC
-31		CGCTCTTGT
-32		CGCTCCTGT
-33		CGCTCATGT
-34		CGCTCGTGT
-35		CGCAGTTGT
-36		CGCAGCTGT

FIG. 46-1 Translation of DNA k9-1

```
     GlyCysProGluArgLeuAlaSerCysArgProLeuThrAspPheAspGlnGlyTrpGly
  1  CAGGCTGTCCTGAGAGGCTAGCCAGCTGCCGACCCCTTACCGATTTTGACCAGGGCTGGG
     GTCCGACAGGACTCTCCGATCGGTCGACGGCTGGGGAATGGCTAAAACTGGTCCCGACCC

     ProIleSerTyrAlaAsnGlySerGlyProAspGlnArgProTyrCysTrpHisTyrPro
 61  GCCCTATCAGTTATGCCAACGGAAGCGGCCCCGACCAGCGCCCCTACTGCTGGCACTACC
     CGGGATAGTCAATACGGTTGCCTTCGCCGGGGCTGGTCGCGGGGATGACGACCGTGATGG

     ProLysProCysGlyIleValProAlaLysSerValCysGlyProValTyrCysPheThr
121  CCCCAAAACCTTGCGGTATTGTGCCCGCGAAGAGTGTGTGTGGTCCGGTATATTGCTTCA
     GGGGTTTTGGAACGCCATAACACGGGCGCTTCTCACACACACCAGGCCATATAACGAAGT

     ProSerProValValValGlyThrThrAspArgSerGlyAlaProThrTyrSerTrpGly
181  CTCCCAGCCCCGTGGTGGTGGGAACGACCGACAGGTCGGGCGCGCCCACCTACAGCTGGG
     GAGGGTCGGGGCACCACCACCCTTGCTGGCTGTCCAGCCCGCGCGGGTGGATGTCGACCC

     GluAsnAspThrAspValPheValLeuAsnAsnThrArgProProLeuGlyAsnTrpPhe
241  GTGAAAATGATACGGACGTCTTCGTCCTTAACAATACCAGGCCACCGCTGGGCAATTGGT
     CACTTTTACTATGCCTGCAGAAGCAGGAATTGTTATGGTCCGGTGGCGACCCGTTAACCA

     GlyCysThrTrpMetAsnSerThrGlyPheThrLysValCysGlyAlaProProCysVal
301  TCGGTTGTACCTGGATGAACTCAACTGGATTCACCAAAGTGTGCGGAGCGCCTCCTTGTG
     AGCCAACATGGACCTACTTGAGTTGACCTAAGTGGTTTCACACGCCTCGCGGAGGAACAC

     IleGlyGlyAlaGlyAsnAsnThrLeuHisCysProThrAspCysPheArgLysHisPro
361  TCATCGGAGGGGCGGGCAACAACACCCTGCACTGCCCCACTGATTGCTTCCGCAAGCATC
     AGTAGCCTCCCCGCCCGTTGTTGTGGGACGTGACGGGGTGACTAACGAAGGCGTTCGTAG

     AspAlaThrTyrSerArgCysGlySerGlyProTrpIleThrProArgCysLeuValAsp
421  CGGACGCCACATACTCTCGGTGCGGCTCCGGTCCCTGGATCACACCCAGGTGCCTGGTCG
     GCCTGCGGTGTATGAGAGCCACGCCGAGGCCAGGGACCTAGTGTGGGTCCACGGACCAGC
                                                     ---------------

     TyrProTyrArgLeuTrpHisTyrProCysThrIleAsnTyrThrIlePheLysIleArg
481  ACTACCCGTATAGGCTTTGGCATTATCCTTGTACCATCAACTACACTATATTTAAAATCA
     TGATGGGCATATCCGAAACCGTAATAGGAACATGGTAGTTGATGTGATATAAATTTTAGT
     ------------------------------------------------------------

     MetTyrValGlyGlyValGluHisArgLeuGluAlaAlaCysAsnTrpThrArgGlyGlu
541  GGATGTACGTGGGAGGGGTCGAGCACAGGCTGGAAGCTGCCTGCAACTGGACGCGGGGCG
     CCTACATGCACCCTCCCCAGCTCGTGTCCGACCTTCGACGGACGTTGACCTGCGCCCCGC
     ------------------------------------------------------------

     ArgCysAspLeuGluAspArgAspArgSerGluLeuSerProLeuLeuLeuThrThrThr
601  AACGTTGCGATCTGGAAGATAGGGACAGGTCCGAGCTCAGCCCGTTACTGCTGACCACTA
     TTGCAACGCTAGACCTTCTATCCCTGTCCAGGCTCGAGTCGGGCAATGACGACTGGTGAT
     ------------------------------------------------------------

     GlnTrpGlnValLeuProCysSerPheThrThrLeuProAlaLeuSerThrGlyLeuIle
661  CACAGTGGCAGGTCCTCCCGTGTTCCTTCACAACCCTGCCAGCCTTGTCCACCGGCCTCA
     GTGTCACCGTCCAGGAGGGCACAAGGAAGTGTTGGGACGGTCGGAACAGGTGGCCGGAGT
     -----Overlap with Combined ORF of DNAs 12f through 15e------
     HisLeuHisGlnAsnIleValAspValGlnTyrLeuTyrGlyValGlySerSerIleAla
721  TCCACCTCCACCAGAACATTGTGGACGTGCAGTACTTGTACGGGGTGGGGTCAAGCATCG
     AGGTGGAGGTGGTCTTGTAACACCTGCACGTCATGAACATGCCCCACCCCAGTTCGTAGC
     ------------------------------------------------------------

     SerTrpAlaIleLysTrpGluTyrValValLeuLeuPheLeuLeuLeuAlaAspAlaArg
781  CGTCCTGGGCCATTAAGTGGGAGTACGTCGTCCTCCTGTTCCTTCTGCTTGCAGACGCGC
     GCAGGACCCGGTAATTCACCCTCATGCAGCAGGAGGACAAGGAAGACGAACGTCTGCGCG
     ------------------------------------------------------------
```

```
        ValCysSerCysLeuTrpMetMetLeuLeuIleSerGlnAlaGluAlaAlaLeuGluAsn
 841   GCGTCTGCTCCTGCTTGTGGATGATGCTACTCATATCCCAAGCGGAAGCGGCTTTGGAGA
       CGCAGACGAGGACGAACACCTACTACGATGAGTATAGGGTTCGCCTTCGCCGAAACCTCT
       ------------------------------------------------------------
        LeuValIleLeuAsnAlaAlaSerLeuAlaGlyThrHisGlyLeuValSerPheLeuVal
 901   ACCTCGTAATACTTAATGCAGCATCCCTGGCCGGGACGCACGGTCTTGTATCCTTCCTCG
       TGGAGCATTATGAATTACGTCGTAGGGACCGGCCCTGCGTGCCAGAACATAGGAAGGAGC
       ------------------------------------------------------------
        PhePheCysPheAlaTrpTyrLeuLysGlyLysTrpValProGlyAlaValTyrThrPhe
 961   TGTTCTTCTGCTTTGCATGGTATCTGAAGGGTAAGTGGGTGCCCGGAGCGGTCTACACCT
       ACAAGAAGACGAAACGTACCATAGACTTCCCATTCACCCACGGGCCTCGCCAGATGTGGA
       ------------------------------------------------------------
        TyrGlyMetTrpProLeuLeuLeuLeuLeuLeuAlaLeuProGlnArgAlaTyrAlaLeu
1021   TCTACGGGATGTGGCCTCTCCTCCTGCTCCTGTTGGCGTTGCCCCAGCGGGCGTACGCGC
       AGATGCCCTACACCGGAGAGGAGGACGAGGACAACCGCAACGGGGTCGCCCGCATGCGCG
       ------------------------------------------------------------
        AspThrGluValAlaAlaSerCysGlyGlyValValLeuValGlyLeuMetAlaLeuThr
1081   TGGACACGGAGGTGGCCGCGTCGTGTGGCGGTGTTGTTCTCGTCGGGTTGATGGCGCTAA
       ACCTGTGCCTCCACCGGCGCAGCACACCGCCACAACAAGAGCAGCCCAACTACCGCGATT
       ------------------------------------------------------------
        LeuSerProTyrTyrLysArgTyrIleSerTrpCysLeuTrpTrpLeuGlnTyrPheLeu
1141   CTCTGTCACCATATTACAAGCGCTATATCAGCTGGTGCTTGTGGTGGCTTCAGTATTTTC
       GAGACAGTGGTATAATGTTCGCGATATAGTCGACCACGAACACCACCGAAGTCATAAAAG
       ------------------------------------------------------------
        ThrArgValGluAlaGlnLeuHisValTrpIleProProLeuAsnValArgGlyGlyArg
1201   TGACCAGAGTGGAAGCGCAACTGCACGTGTGGATTCCCCCCCTCAACGTCCGAGGGGGGC
       ACTGGTCTCACCTTCGCGTTGACGTGCACACCTAAGGGGGGGAGTTGCAGGCTCCCCCCG
       ------------------------------------------------------------
        AspAlaValIleLeuLeuMetCysAlaValHisProThrLeuValPheAspIleThrLys
1261   GCGACGCTGTCATCTTACTCATGTGTGCTGTACACCCGACTCTGGTATTTGACATCACCA
       CGCTGCGACAGTAGAATGAGTACACACGACATGTGGGCTGAGACCATAAACTGTAGTGGT
       ---------------------------------------------
        LeuLeuLeuAlaValPheGlyProLeuTrpIleLeuGlnAla
1321   AATTGCTGCTGGCCGTCTTCGGACCCCTTTGGATTCTTCAAGCCAG
       TTAACGACGACCGGCAGAAGCCTGGGGAAACCTAAGAAGTTCGGTC
```

FIG. 46-2

FIG. 47-1 COMBINED ORF OF DNAs K9-1 through 15e

```
     GlyCysProGluArgLeuAlaSerCysArgProLeuThrAspPheAspGlnGlyTrpGly
  1  CAGGCTGTCCTGAGAGGCTAGCCAGCTGCCGACCCCTTACCGATTTTGACCAGGGCTGGG
     GTCCGACAGGACTCTCCGATCGGTCGACGGCTGGGGAATGGCTAAAACTGGTCCCGACCC

     ProIleSerTyrAlaAsnGlySerGlyProAspGlnArgProTyrCysTrpHisTyrPro
 61  GCCCTATCAGTTATGCCAACGGAAGCGGCCCCGACCAGCGCCCCTACTGCTGGCACTACC
     CGGGATAGTCAATACGGTTGCCTTCGCCGGGGCTGGTCGCGGGGATGACGACCGTGATGG

     ProLysProCysGlyIleValProAlaLysSerValCysGlyProValTyrCysPheThr
121  CCCCAAAACCTTGCGGTATTGTGCCCGCGAAGAGTGTGTGTGGTCCGGTATATTGCTTCA
     GGGGTTTTGGAACGCCATAACACGGGCGCTTCTCACACACACCAGGCCATATAACGAAGT

     ProSerProValValValGlyThrThrAspArgSerGlyAlaProThrTyrSerTrpGly
181  CTCCCAGCCCCGTGGTGGTGGGAACGACCGACAGGTCGGGCGCGCCCACCTACAGCTGGG
     GAGGGTCGGGGCACCACCACCCTTGCTGGCTGTCCAGCCCGCGCGGGTGGATGTCGACCC

     GluAsnAspThrAspValPheValLeuAsnAsnThrArgProProLeuGlyAsnTrpPhe
241  GTGAAAATGATACGGACGTCTTCGTCCTTAACAATACCAGGCCACCGCTGGGCAATTGGT
     CACTTTTACTATGCCTGCAGAAGCAGGAATTGTTATGGTCCGGTGGCGACCCGTTAACCA

     GlyCysThrTrpMetAsnSerThrGlyPheThrLysValCysGlyAlaProProCysVal
301  TCGGTTGTACCTGGATGAACTCAACTGGATTCACCAAAGTGTGCGGAGCGCCTCCTTGTG
     AGCCAACATGGACCTACTTGAGTTGACCTAAGTGGTTTCACACGCCTCGCGGAGGAACAC

     IleGlyGlyAlaGlyAsnAsnThrLeuHisCysProThrAspCysPheArgLysHisPro
361  TCATCGGAGGGGCGGGCAACAACACCCTGCACTGCCCCACTGATTGCTTCCGCAAGCATC
     AGTAGCCTCCCCGCCCGTTGTTGTGGGACGTGACGGGGTGACTAACGAAGGCGTTCGTAG

     AspAlaThrTyrSerArgCysGlySerGlyProTrpIleThrProArgCysLeuValAsp
421  CGGACGCCACATACTCTCGGTGCGGCTCCGGTCCCTGGATCACACCCAGGTGCCTGGTCG
     GCCTGCGGTGTATGAGAGCCACGCCGAGGCCAGGGACCTAGTGTGGGTCCACGGACCAGC

     TyrProTyrArgLeuTrpHisTyrProCysThrIleAsnTyrThrIlePheLysIleArg
481  ACTACCCGTATAGGCTTTGGCATTATCCTTGTACCATCAACTACACCATATTTAAAATCA
     TGATGGGCATATCCGAAACCGTAATAGGAACATGGTAGTTGATGTGGTATAAATTTTAGT

     MetTyrValGlyGlyValGluHisArgLeuGluAlaAlaCysAsnTrpThrArgGlyGlu
541  GGATGTACGTGGGAGGGGTCGAACACAGGCTGGAAGCTGCCTGCAACTGGACGCGGGGCG
     CCTACATGCACCCTCCCCAGCTTGTGTCCGACCTTCGACGGACGTTGACCTGCGCCCCGC

     ArgCysAspLeuGluAspArgAspArgSerGluLeuSerProLeuLeuLeuThrThrThr
601  AACGTTGCGATCTGGAAGACAGGGACAGGTCCGAGCTCAGCCCGTTACTGCTGACCACTA
     TTGCAACGCTAGACCTTCTGTCCCTGTCCAGGCTCGAGTCGGGCAATGACGACTGGTGAT

     GlnTrpGlnValLeuProCysSerPheThrThrLeuProAlaLeuSerThrGlyLeuIle
661  CACAGTGGCAGGTCCTCCCGTGTTCCTTCACAACCCTACCAGCCTTGTCCACCGGCCTCA
     GTGTCACCGTCCAGGAGGGCACAAGGAAGTGTTGGGATGGTCGGAACAGGTGGCCGGAGT

     HisLeuHisGlnAsnIleValAspValGlnTyrLeuTyrGlyValGlySerSerIleAla
721  TCCACCTCCACCAGAACATTGTGGACGTGCAGTACTTGTACGGGGTGGGGTCAAGCATCG
     AGGTGGAGGTGGTCTTGTAACACCTGCACGTCATGAACATGCCCCACCCCAGTTCGTAGC

     SerTrpAlaIleLysTrpGluTyrValValLeuLeuPheLeuLeuLeuAlaAspAlaArg
781  CGTCCTGGGCCATTAAGTGGGAGTACGTCGTTCTCCTGTTCCTTCTGCTTGCAGACGCGC
     GCAGGACCCGGTAATTCACCCTCATGCAGCAAGAGGACAAGGAAGACGAACGTCTGCGCG

     ValCysSerCysLeuTrpMetMetLeuLeuIleSerGlnAlaGluAlaAlaLeuGluAsn
841  GCGTCTGCTCCTGCTTGTGGATGATGCTACTCATATCCCAAGCGGAGGCGGCTTTGGAGA
     CGCAGACGAGGACGAACACCTACTACGATGAGTATAGGGTTCGCCTCCGCCGAAACCTCT

     LeuValIleLeuAsnAlaAlaSerLeuAlaGlyThrHisGlyLeuValSerPheLeuVal
901  ACCTCGTAATACTTAATGCAGCATCCCTGGCCGGGACGCACGGTCTTGTATCCTTCCTCG
     TGGAGCATTATGAATTACGTCGTAGGGACCGGCCCTGCGTGCCAGAACATAGGAAGGAGC
```

```
      PhePheCysPheAlaTrpTyrLeuLysGlyLysTrpValProGlyAlaValTyrThrPhe
 961  TGTTCTTCTGCTTTGCATGGTATTTGAAGGGTAAGTGGGTGCCCGGAGCGGTCTACACCT
      ACAAGAAGACGAAACGTACCATAAACTTCCCATTCACCCACGGGCCTCGCCAGATGTGGA

      TyrGlyMetTrpProLeuLeuLeuLeuLeuLeuAlaLeuProGlnArgAlaTyrAlaLeu
1021  TCTACGGGATGTGGCCTCTCCTCCTGCTCCTGTTGGCGTTGCCCCAGCGGGCGTACGCGC
      AGATGCCCTACACCGGAGAGGAGGACGAGGACAACCGCAACGGGGTCGCCCGCATGCGCG

      AspThrGluValAlaAlaSerCysGlyGlyValValLeuValGlyLeuMetAlaLeuThr
1081  TGGACACGGAGGTGGCCGCGTCGTGTGGCGGTGTTGTTCTCGTCGGGTTGATGGCGCTGA
      ACCTGTGCCTCCACCGGCGCAGCACACCGCCACAACAAGAGCAGCCCAACTACCGCGACT

      LeuSerProTyrTyrLysArgTyrIleSerTrpCysLeuTrpTrpLeuGlnTyrPheLeu
1141  CTCTGTCACCATATTACAAGCGCTATATCAGCTGGTGCTTGTGGTGGCTTCAGTATTTTC
      GAGACAGTGGTATAATGTTCGCGATATAGTCGACCACGAACACCACCGAAGTCATAAAAG

      ThrArgValGluAlaGlnLeuHisValTrpIleProProLeuAsnValArgGlyGlyArg
1201  TGACCAGAGTGGAAGCGCAACTGCACGTGTGGATTCCCCCCCTCAACGTCCGAGGGGGGC
      ACTGGTCTCACCTTCGCGTTGACGTGCACACCTAAGGGGGGGAGTTGCAGGCTCCCCCCG

      AspAlaValIleLeuLeuMetCysAlaValHisProThrLeuValPheAspIleThrLys
1261  GCGACGCCGTCATCTTACTCATGTGTGCTGTACACCCGACTCTGGTATTTGACATCACCA
      CGCTGCGGCAGTAGAATGAGTACACACGACATGTGGGCTGAGACCATAAACTGTAGTGGT

      LeuLeuLeuAlaValPheGlyProLeuTrpIleLeuGlnAlaSerLeuLeuLysValPro
1321  AATTGCTGCTGGCCGTCTTCGGACCCCTTTGGATTCTTCAAGCCAGTTTGCTTAAAGTAC
      TTAACGACGACCGGCAGAAGCCTGGGGAAACCTAAGAAGTTCGGTCAAACGAATTTCATG

      TyrPheValArgValGlnGlyLeuLeuArgPheCysAlaLeuAlaArgLysMetIleGly
1381  CCTACTTTGTGCGCGTCCAAGGCCTTCTCCGGTTCTGCGCGTTAGCGCGGAAGATGATCG
      GGATGAAACACGCGCAGGTTCCGGAAGAGGCCAAGACGCGCAATCGCGCCTTCTACTAGC

      GlyHisTyrValGlnMetValIleIleLysLeuGlyAlaLeuThrGlyThrTyrValTyr
1441  GAGGCCATTACGTGCAAATGGTCATCATTAAGTTAGGGGCGCTTACTGGCACCTATGTTT
      CTCCGGTAATGCACGTTTACCAGTAGTAATTCAATCCCCGCGAATGACCGTGGATACAAA

      AsnHisLeuThrProLeuArgAspTrpAlaHisAsnGlyLeuArgAspLeuAlaValAla
1501  ATAACCATCTCACTCCTCTTCGGGACTGGGCGCACAACGGCTTGCGAGATCTGGCCGTGG
      TATTGGTAGAGTGAGGAGAAGCCCTGACCCGCGTGTTGCCGAACGCTCTAGACCGGCACC

      ValGluProValValPheSerGlnMetGluThrLysLeuIleThrTrpGlyAlaAspThr
1561  CTGTAGAGCCAGTCGTCTTCTCCCAAATGGAGACCAAGCTCATCACGTGGGGGGCAGATA
      GACATCTCGGTCAGCAGAAGAGGGTTTACCTCTGGTTCGAGTAGTGCACCCCCCGTCTAT

      AlaAlaCysGlyAspIleIleAsnGlyLeuProValSerAlaArgArgGlyArgGluIle
1621  CCGCCGCGTGCGGTGACATCATCAACGGCTTGCCTGTTTCCGCCCGCAGGGGCCGGGAGA
      GGCGGCGCACGCCACTGTAGTAGTTGCCGAACGGACAAAGGCGGGCGTCCCCGGCCCTCT

      LeuLeuGlyProAlaAspGlyMetValSerLysGlyTrpArgLeuLeuAlaProIleThr
1681  TACTGCTCGGGCCAGCCGATGGAATGGTCTCCAAGGGGTGGAGGTTGCTGGCGCCCATCA
      ATGACGAGCCCGGTCGGCTACCTTACCAGAGGTTCCCCACCTCCAACGACCGCGGGTAGT

      AlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThrSerLeuThrGlyArg
1741  CGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACCAGCCTAACTGGCC
      GCCGCATGCGGGTCGTCTGTTCCCCGGAGGATCCCACGTATTAGTGGTCGGATTGACCGG

      AspLysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeu
1801  GGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCTGCCCAAACCTTCC
      CCCTGTTTTTGGTTCACCTCCCACTCCAGGTCTAACACAGTTGACGACGGGTTTGGAAGG

      AlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIle
1861  TGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCCGGAACGAGGACCA
      ACCGTTGCACGTAGTTACCCCACACGACCTGACAGATGGTGCCCCGGCCTTGCTCCTGGT

      AlaSerProLysGlyProValIleGlnMetTyrThrAsnValAspGlnAspLeuValGly
```

FIG. 47-2

1921 TCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGACCAAGACCTTGTGG
AGCGCAGTGGGTTCCCAGGACAGTAGGTCTACATATGGTTACATCTGGTTCTGGAACACC

TrpProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeu
1981 GCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGCGGCTCCTCGGACC
CGACCGGGCGAGGCGTTCCATCGGCGAGTAACTGTGGGACGTGAACGCCGAGGAGCCTGG

TyrLeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGly
2041 TTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGGGGTGATAGCAGGG
AAATGGACCAGTGCTCCGTGCGGCTACAGTAAGGGCACGCGGCCGCCCCACTATCGTCCC

SerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeu
2101 GCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCGGGGGGTCCGCTGT
CGTCGGACGACAGCGGGGCCGGGTAAAGGATGAACTTTCCGAGGAGCCCCCCAGGCGACA

CysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAla
2161 TGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGCACCCGTGGAGTGG
ACACGGGGCGCCCCGTGCGGCACCCGTATAAATCCCGGCGCCACACGTGGGCACCTCACC

LysAlaValAspPheIleProValGluAsnLeuGluThrThrMetArgSerProValPhe
2221 CTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATGAGGTCCCCGGTGT
GATTCCGCCACCTGAAATAGGGACACCTCTTGGATCTCTGTTGGTACTCCAGGGGCCACA

ThrAspAsnSerSerProProValValProGlnSerPheGlnValAlaHisLeuHisAla
2281 TCACGGATAACTCCTCTCCACCAGTAGTGCCCCAGAGCTTCCAGGTGGCTCACCTCCATG
AGTGCCTATTGAGGAGAGGTGGTCATCACGGGGTCTCGAAGGTCCACCGAGTGGAGGTAC

ProThrGlySerGlyLysSerThrLysValProAlaAlaTyrAlaAlaGlnGlyTyrLys
2341 CTCCCACAGGCAGCGGCAAAAGCACCAAGGTCCCGGCTGCATATGCAGCTCAGGGCTATA
GAGGGTGTCCGTCGCCGTTTTCGTGGTTCCAGGGCCGACGTATACGTCGAGTCCCGATAT

ValLeuValLeuAsnProSerValAlaAlaThrLeuGlyPheGlyAlaTyrMetSerLys
2401 AGGTGCTAGTACTCAACCCCTCTGTTGCTGCAACACTGGGCTTTGGTGCTTACATGTCCA
TCCACGATCATGAGTTGGGGAGACAACGACGTTGTGACCCGAAACCACGAATGTACAGGT

AlaHisGlyIleAspProAsnIleArgThrGlyValArgThrIleThrThrGlySerPro
2461 AGGCTCATGGGATCGATCCTAACATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCC
TCCGAGTACCCTAGCTAGGATTGTAGTCCTGGCCCCACTCTTGTTAATGGTGACCGTCGG

IleThrTyrSerThrTyrGlyLysPheLeuAlaAspGlyGlyCysSerGlyGlyAlaTyr
2521 CCATCACGTACTCCACCTACGGCAAGTTCCTTGCCGACGGCGGGTGCTCGGGGGGCGCTT
GGTAGTGCATGAGGTGGATGCCGTTCAAGGAACGGCTGCCGCCCACGAGCCCCCCGCGAA

AspIleIleIleCysAspGluCysHisSerThrAspAlaThrSerIleLeuGlyIleGly
2581 ATGACATAATAATTTGTGACGAGTGCCACTCCACGGATGCCACATCCATCTTGGGCATCG
TACTGTATTATTAAACACTGCTCACGGTGAGGTGCCTACGGTGTAGGTAGAACCCGTAGC

ThrValLeuAspGlnAlaGluThrAlaGlyAlaArgLeuValValLeuAlaThrAlaThr
2641 GCACTGTCCTTGACCAAGCAGAGACTGCGGGGGCGAGACTGGTTGTGCTCGCCACCGCCA
CGTGACAGGAACTGGTTCGTCTCTGACGCCCCCGCTCTGACCAACACGAGCGGTGGCGGT

ProProGlySerValThrValProHisProAsnIleGluGluValAlaLeuSerThrThr
2701 CCCCTCCGGGCTCCGTCACTGTGCCCCATCCCAACATCGAGGAGGTTGCTCTGTCCACCA
GGGGAGGCCCGAGGCAGTGACACGGGGTAGGGTTGTAGCTCCTCCAACGAGACAGGTGGT

GlyGluIleProPheTyrGlyLysAlaIleProLeuGluValIleLysGlyGlyArgHis
2761 CCGGAGAGATCCCTTTTTACGGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGGAGAC
GGCCTCTCTAGGGAAAAATGCCGTTCCGATAGGGGGAGCTTCATTAGTTCCCCCCCTCTG

LeuIlePheCysHisSerLysLysLysCysAspGluLeuAlaAlaLysLeuValAlaLeu
2821 ATCTCATCTTCTGTCATTCAAAGAAGAAGTGCGACGAACTCGCCGCAAAGCTGGTCGCAT
TAGAGTAGAAGACAGTAAGTTTCTTCTTCACGCTGCTTGAGCGGCGTTTCGACCAGCGTA

GlyIleAsnAlaValAlaTyrTyrArgGlyLeuAspValSerValIleProThrSerGly
2881 TGGGCATCAATGCCGTGGCCTACTACCGCGGTCTTGACGTGTCCGTCATCCCGACCAGCG
ACCCGTAGTTACGGCACCGGATGATGGCGCCAGAACTGCACAGGCAGTAGGGCTGGTCGC

FIG. 47-3

```
      AspValValValValAlaThrAspAlaLeuMetThrGlyTyrThrGlyAspPheAspSer
2941  GCGATGTTGTCGTCGTGGCAACCGATGCCCTCATGACCGGCTATACCGGCGACTTCGACT
      CGCTACAACAGCAGCACCGTTGGCTACGGGAGTACTGGCCGATATGGCCGCTGAAGCTGA

      ValIleAspCysAsnThrCysValThrGlnThrValAspPheSerLeuAspProThrPhe
3001  CGGTGATAGACTGCAATACGTGTGTCACCCAGACAGTCGATTTCAGCCTTGACCCTACCT
      GCCACTATCTGACGTTATGCACACAGTGGGTCTGTCAGCTAAAGTCGGAACTGGGATGGA

      ThrIleGluThrIleThrLeuProGlnAspAlaValSerArgThrGlnArgArgGlyArg
3061  TCACCATTGAGACAATCACGCTCCCCCAGGATGCTGTCTCCCGCACTCAACGTCGGGGCA
      AGTGGTAACTCTGTTAGTGCGAGGGGGTCCTACGACAGAGGGCGTGAGTTGCAGCCCCGT

      ThrGlyArgGlyLysProGlyIleTyrArgPheValAlaProGlyGluArgProSerGly
3121  GGACTGGCAGGGGGAAGCCAGGCATCTACAGATTTGTGGCACCGGGGGAGCGCCCCTCCG
      CCTGACCGTCCCCCTTCGGTCCGTAGATGTCTAAACACCGTGGCCCCCTCGCGGGGAGGC

      MetPheAspSerSerValLeuCysGluCysTyrAspAlaGlyCysAlaTrpTyrGluLeu
3181  GCATGTTCGACTCGTCCGTCCTCTGTGAGTGCTATGACGCAGGCTGTGCTTGGTATGAGC
      CGTACAAGCTGAGCAGGCAGGAGACACTCACGATACTGCGTCCGACACGAACCATACTCG

      ThrProAlaGluThrThrValArgLeuArgAlaTyrMetAsnThrProGlyLeuProVal
3241  TCACGCCCGCCGAGACTACAGTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCG
      AGTGCGGGCGGCTCTGATGTCAATCCGATGCTCGCATGTACTTGTGGGGCCCCGAAGGGC

      CysGlnAspHisLeuGluPheTrpGluGlyValPheThrGlyLeuThrHisIleAspAla
3301  TGTGCCAGGACCATCTTGAATTTTGGGAGGGCGTCTTTACAGGCCTCACTCATATAGATG
      ACACGGTCCTGGTAGAACTTAAAACCCTCCCGCAGAAATGTCCGGAGTGAGTATATCTAC

      HisPheLeuSerGlnThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGln
3361  CCCACTTTCTATCCCAGACAAAGCAGAGTGGGGAGAACCTTCCTTACCTGGTAGCGTACC
      GGGTGAAAGATAGGGTCTGTTTCGTCTCACCCCTCTTGGAAGGAATGGACCATCGCATGG

      AlaThrValCysAlaArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLysCys
3421  AAGCCACCGTGTGCGCTAGGGCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGT
      TTCGGTGGCACACGCGATCCCGAGTTCGGGGAGGGGGTAGCACCCTGGTCTACACCTTCA

      LeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAla
3481  GTTTGATTCGCCTCAAGCCCACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCG
      CAAACTAAGCGGAGTTCGGGTGGGAGGTACCCGGTTGTGGGGACGATATGTCTGACCCGC

      ValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIleMetThrCysMetSer
3541  CTGTTCAGAATGAAATCACCCTGACGCACCCAGTCACCAAATACATCATGACATGCATGT
      GACAAGTCTTACTTTAGTGGGACTGCGTGGGTCAGTGGTTTATGTAGTACTGTACGTACA

      AlaAspLeuGluValValThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeu
3601  CGGCCGACCTGGAGGTCGTCACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTT
      GCCGGCTGGACCTCCAGCAGTGCTCGTGGACCCACGAGCAACCGCCGCAGGACCGACGAA

      AlaAlaTyrCysLeuSerThrGlyCysValValIleValGlyArgValValLeuSerGly
3661  TGGCCGCGTATTGCCTGTCAACAGGCTGCGTGGTCATAGTGGGCAGGGTCGTCTTGTCCG
      ACCGGCGCATAACGGACAGTTGTCCGACGCACCAGTATCACCCGTCCCAGCAGAACAGGC

      LysProAlaIleIleProAspArgGluValLeuTyrArgGluPheAspGluMetGluGlu
3721  GGAAGCCGGCAATCATACCTGACAGGGAAGTCCTCTACCGAGAGTTCGATGAGATGGAAG
      CCTTCGGCCGTTAGTATGGACTGTCCCTTCAGGAGATGGCTCTCAAGCTACTCTACCTTC

      CysSerGlnHisLeuProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPheLysGln
3781  AGTGCTCTCAGCACTTACCGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGC
      TCACGAGAGTCGTGAATGGCATGTAGCTCGTTCCCTACTACGAGCGGCTCGTCAAGTTCG

      LysAlaLeuGlyLeuLeuGlnThrAlaSerArgGlnAlaGluValIleAlaProAlaVal
3841  AGAAGGCCCTCGGCCTCCTGCAGACCGCGTCCCGTCAGGCAGAGGTTATCGCCCCTGCTG
      TCTTCCGGGAGCCGGAGGACGTCTGGCGCAGGGCAGTCCGTCTCCAATAGCGGGGACGAC

      GlnThrAsnTrpGlnLysLeuGluThrPheTrpAlaLysHisMetTrpAsnPheIleSer
```

FIG. 47-4

3901 TCCAGACCAACTGGCAAAAACTCGAGACCTTCTGGGCGAAGCATATGTGGAACTTCATCA
AGGTCTGGTTGACCGTTTTTGAGCTCTGGAAGACCCGCTTCGTATACACCTTGAAGTAGT

GlyIleGlnTyrLeuAlaGlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeu
3961 GTGGGATACAATACTTGGCGGGCTTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCAT
CACCCTATGTTATGAACCGCCCGAACAGTTGCGACGGACCATTGGGGCGGTAACGAAGTA

MetAlaPheThrAlaAlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsn
4021 TGATGGCTTTTACAGCTGCTGTCACCAGCCCACTAACCACTAGCCAAACCCTCCTCTTCA
ACTACCGAAAATGTCGACGACAGTGGTCGGGTGATTGGTGATCGGTTTGGGAGGAGAAGT

IleLeuGlyGlyTrpValAlaAlaGlnLeuAlaAlaProGlyAlaAlaThrAlaPheVal
4081 ACATATTGGGGGGGTGGGTGGCTGCCCAGCTCGCCGCCCCCGGTGCCGCTACTGCCTTTG
TGTATAACCCCCCCACCCACCGACGGGTCGAGCGGCGGGGGCCACGGCGATGACGGAAAC

GlyAlaGlyLeuAlaGlyAlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAsp
4141 TGGGCGCTGGCTTAGCTGGCGCCGCCATCGGCAGTGTTGGACTGGGGAAGGTCCTCATAG
ACCCGCGACCGAATCGACCGCGGCGGTAGCCGTCACAACCTGACCCCTTCCAGGAGTATC

IleLeuAlaGlyTyrGlyAlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSer
4201 ACATCCTTGCAGGGTATGGCGCGGGCGTGGCGGGAGCTCTTGTGGCATTCAAGATCATGA
TGTAGGAACGTCCCATACCGCGCCCGCACCGCCCTCGAGAACACCGTAAGTTCTAGTACT

GlyGluValProSerThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGly
4261 GCGGTGAGGTCCCCTCCACGGAGGACCTGGTCAATCTACTGCCCGCCATCCTCTCGCCCG
CGCCACTCCAGGGGAGGTGCCTCCTGGACCAGTTAGATGACGGGCGGTAGGAGAGCGGGC

AlaLeuValValGlyValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGlu
4321 GAGCCCTCGTAGTCGGCGTGGTCTGTGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCG
CTCGGGAGCATCAGCCGCACCAGACACGTCGTTATGACGCGGCCGTGCAACCGGGCCCGC

GlyAlaValGlnTrpMetAsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSer
4381 AGGGGGCAGTGCAGTGGATGAACCGGCTGATAGCCTTCGCCTCCCGGGGGAACCATGTTT
TCCCCCGTCACGTCACCTACTTGGCCGACTATCGGAAGCGGAGGGCCCCCTTGGTACAAA

ProThrHisTyrValProGluSerAspAlaAlaAlaArgValThrAlaIleLeuSerSer
4441 CCCCCACGCACTACGTGCCGGAGAGCGATGCAGCTGCCCGCGTCACTGCCATACTCAGCA
GGGGGTGCGTGATGCACGGCCTCTCGCTACGTCGACGGGCGCAGTGACGGTATGAGTCGT

LeuThrValThrGlnLeuLeuArgArgLeuHisGlnTrpIleSerSerGluCysThrThr
4501 GCCTCACTGTAACCCAGCTCCTGAGGCGACTGCACCAGTGGATAAGCTCGGAGTGTACCA
CGGAGTGACATTGGGTCGAGGACTCCGCTGACGTGGTCACCTATTCGAGCCTCACATGGT

ProCysSerGlySerTrpLeuArgAspIleTrpAspTrpIleCysGluValLeuSerAsp
4561 CTCCATGCTCCGGTTCCTGGCTAAGGGACATCTGGGACTGGATATGCGAGGTGTTGAGCG
GAGGTACGAGGCCAAGGACCGATTCCCTGTAGACCCTGACCTATACGCTCCACAACTCGC

PheLysThrTrpLeuLysAlaLysLeuMetProGlnLeuProGlyIleProPheValSer
4621 ACTTTAAGACCTGGCTAAAAGCTAAGCTCATGCCACAGCTGCCTGGGATCCCCTTTGTGT
TGAAATTCTGGACCGATTTTCGATTCGAGTACGGTGTCGACGGACCCTAGGGGAAACACA

CysGlnArgGlyTyrLysGlyValTrpArgValAspGlyIleMetHisThrArgCysHis
4681 CCTGCCAGCGCGGGTATAAGGGGGTCTGGCGAGTGGACGGCATCATGCACACTCGCTGCC
GGACGGTCGCGCCCATATTCCCCCAGACCGCTCACCTGCCGTAGTACGTGTGAGCGACGG

CysGlyAlaGluIleThrGlyHisValLysAsnGlyThrMetArgIleValGlyProArg
4741 ACTGTGGAGCTGAGATCACTGGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTA
TGACACCTCGACTCTAGTGACCTGTACAGTTTTTGCCCTGCTACTCCTAGCAGCCAGGAT

ThrCysArgAsnMetTrpSerGlyThrPheProIleAsnAlaTyrThrThrGlyProCys
4801 GGACCTGCAGGAACATGTGGAGTGGGACCTTCCCCATTAATGCCTACACCACGGGCCCCT
CCTGGACGTCCTTGTACACCTCACCCTGGAAGGGGTAATTACGGATGTGGTGCCCGGGGA

ThrProLeuProAlaProAsnTyrThrPheAlaLeuTrpArgValSerAlaGluGluTyr
4861 GTACCCCCCCTTCCTGCGCCGAACTACACGTTCGCGCTATGGAGGGTGTCTGCAGAGGAAT
CATGGGGGGAAGGACGCGGCTTGATGTGCAAGCGCGATACCTCCCACAGACGTCTCCTTA

FIG. 47-5

```
      ValGluIleArgGlnValGlyAspPheHisTyrValThrGlyMetThrThrAspAsnLeu
4921  ATGTGGAGATAAGGCAGGTGGGGGACTTCCACTACGTGACGGGTATGACTACTGACAATC
      TACACCTCTATTCCGTCCACCCCCTGAAGGTGATGCACTGCCCATACTGATGACTGTTAG

      LysCysProCysGlnValProSerProGluPhePheThrGluLeuAspGlyValArgLeu
4981  TCAAATGCCCGTGCCAGGTCCCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCC
      AGTTTACGGGCACGGTCCAGGGTAGCGGGCTTAAAAAGTGTCTTAACCTGCCCCACGCGG

      HisArgPheAlaProProCysLysProLeuLeuArgGluGluValSerPheArgValGly
5041  TACATAGGTTTGCGCCCCCCTGCAAGCCCTTGCTGCGGGAGGAGGTATCATTCAGAGTAG
      ATGTATCCAAACGCGGGGGGACGTTCGGGAACGACGCCCTCCTCCATAGTAAGTCTCATC

      LeuHisGluTyrProValGlySerGlnLeuProCysGluProGluProAspValAlaVal
5101  GACTCCACGAATACCCGGTAGGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTGGCCG
      CTGAGGTGCTTATGGGCCATCCCAGCGTTAATGGAACGCTCGGGCTTGGCCTGCACCGGC

      LeuThrSerMetLeuThrAspProSerHisIleThrAlaGluAlaAlaGlyArgArgLeu
5161  TGTTGACGTCCATGCTCACTGATCCCTCCCATATAACAGCAGAGGCGGCCGGGCGAAGGT
      ACAACTGCAGGTACGAGTGACTAGGGAGGGTATATTGTCGTCTCCGCCGGCCCGCTTCCA

      AlaArgGlySerProProSerValAlaSerSerSerAlaSerGlnLeuSerAlaProSer
5221  TGGCGAGGGGATCACCCCCCTCTGTGGCCAGCTCCTCGGCTAGCCAGCTATCCGCTCCAT
      ACCGCTCCCCTAGTGGGGGGAGACACCGGTCGAGGAGCCGATCGGTCGATAGGCGAGGTA

      LeuLysAlaThrCysThrAlaAsnHisAspSerProAspAlaGluLeuIleGluAlaAsn
5281  CTCTCAAGGCAACTTGCACCGCTAACCATGACTCCCCTGATGCTGAGCTCATAGAGGCCA
      GAGAGTTCCGTTGAACGTGGCGATTGGTACTGAGGGGACTACGACTCGAGTATCTCCGGT

      LeuLeuTrpArgGlnGluMetGlyGlyAsnIleThrArgValGluSerGluAsnLysVal
5341  ACCTCCTATGGAGGCAGGAGATGGGCGGCAACATCACCAGGGTTGAGTCAGAAAACAAAG
      TGGAGGATACCTCCGTCCTCTACCCGCCGTTGTAGTGGTCCCAACTCAGTCTTTTGTTTC

      ValIleLeuAspSerPheAspProLeuValAlaGluGluAspGluArgGluIleSerVal
5401  TGGTGATTCTGGACTCCTTCGATCCGCTTGTGGCGGAGGAGGACGAGCGGGAGATCTCCG
      ACCACTAAGACCTGAGGAAGCTAGGCGAACACCGCCTCCTCCTGCTCGCCCTCTAGAGGC

      ProAlaGluIleLeuArgLysSerArgArgPheAlaGlnAlaLeuProValTrpAlaArg
5461  TACCCGCAGAAATCCTGCGGAAGTCTCGGAGATTCGCCCAGGCCCTGCCCGTTTGGGCGC
      ATGGGCGTCTTTAGGACGCCTTCAGAGCCTCTAAGCGGGTCCGGGACGGGCAAACCCGCG

      ProAspTyrAsnProProLeuValGluThrTrpLysLysProAspTyrGluProProVal
5521  GGCCGGACTATAACCCCCCGCTAGTGGAGACGTGGAAAAAGCCCGACTACGAACCACCTG
      CCGGCCTGATATTGGGGGGCGATCACCTCTGCACCTTTTTCGGGCTGATGCTTGGTGGAC

      ValHisGlyCysProLeuProProProLysSerProProValProProProArgLysLys
5581  TGGTCCATGGCTGTCCGCTTCCACCTCCAAAGTCCCCTCCTGTGCCTCCGCCTCGGAAGA
      ACCAGGTACCGACAGGCGAAGGTGGAGGTTTCAGGGGAGGACACGGAGGCGGAGCCTTCT

      ArgThrValValLeuThrGluSerThrLeuSerThrAlaLeuAlaGluLeuAlaThrArg
5641  AGCGGACGGTGGTCCTCACTGAATCAACCCTATCTACTGCCTTGGCCGAGCTCGCCACCA
      TCGCCTGCCACCAGGAGTGACTTAGTTGGGATAGATGACGGAACCGGCTCGAGCGGTGGT

      SerPheGlySerSerSerThrSerGlyIleThrGlyAspAsnThrThrThrSerSerGlu
5701  GAAGCTTTGGCAGCTCCTCAACTTCCGGCATTACGGGCGACAATACGACAACATCCTCTG
      CTTCGAAACCGTCGAGGAGTTGAAGGCCGTAATGCCCGCTGTTATGCTGTTGTAGGAGAC

      ProAlaProSerGlyCysProProAspSerAspAlaGluSerTyrSerSerMetProPro
5761  AGCCCGCCCCTTCTGGCTGCCCCCCCGACTCCGACGCTGAGTCCTATTCCTCCATGCCCC
      TCGGGCGGGGAAGACCGACGGGGGGGCTGAGGCTGCGACTCAGGATAAGGAGGTACGGGG

      LeuGluGlyGluProGlyAspProAspLeuSerAspGlySerTrpSerThrValSerSer
5821  CCCTGGAGGGGGAGCCTGGGGATCCGGATCTTAGCGACGGGTCATGGTCAACGGTCAGTA
      GGGACCTCCCCCTCGGACCCCTAGGCCTAGAATCGCTGCCCAGTACCAGTTGCCAGTCAT

      GluAlaAsnAlaGluAspValValCysCysSerMetSerTyrSerTrpThrGlyAlaLeu
```

FIG. 47-6

62/63

5881 GTGAGGCCAACGCGGAGGATGTCGTGTGCTGCTCAATGTCTTACTCTTGGACAGGCGCAC
CACTCCGGTTGCGCCTCCTACAGCACACGACGAGTTACAGAATGAGAACCTGTCCGCGTG

ValThrProCysAlaAlaGluGluGlnLysLeuProIleAsnAlaLeuSerAsnSerLeu
5941 TCGTCACCCCGTGCGCCGCGGAAGAACAGAAACTGCCCATCAATGCACTAAGCAACTCGT
AGCAGTGGGGCACGCGGCGCCTTCTTGTCTTTGACGGGTAGTTACGTGATTCGTTGAGCA

LeuArgHisHisAsnLeuValTyrSerThrThrSerArgSerAlaCysGlnArgGlnLys
6001 TGCTACGTCACCACAATTTGGTGTATTCCACCACCTCACGCAGTGCTTGCCAAAGGCAGA
ACGATGCAGTGGTGTTAAACCACATAAGGTGGTGGAGTGCGTCACGAACGGTTTCCGTCT

LysValThrPheAspArgLeuGlnValLeuAspSerHisTyrGlnAspValLeuLysGlu
6061 AGAAAGTCACATTTGACAGACTGCAAGTTCTGGACAGCCATTACCAGGACGTACTCAAGG
TCTTTCAGTGTAAACTGTCTGACGTTCAAGACCTGTCGGTAATGGTCCTGCATGAGTTCC

ValLysAlaAlaAlaSerLysValLysAlaAsnLeuLeuSerValGluGluAlaCysSer
6121 AGGTTAAAGCAGCGGCGTCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGCA
TCCAATTTCGTCGCCGCAGTTTTCACTTCCGATTGAACGATAGGCATCTCCTTCGAACGT

LeuThrProProHisSerAlaLysSerLysPheGlyTyrGlyAlaLysAspValArgCys
6181 GCCTGACGCCCCCACACTCAGCCAAATCCAAGTTTGGTTATGGGGCAAAAGACGTCCGTT
CGGACTGCGGGGGTGTGAGTCGGTTTAGGTTCAAACCAATACCCCGTTTTCTGCAGGCAA

HisAlaArgLysAlaValThrHisIleAsnSerValTrpLysAspLeuLeuGluAspAsn
6241 GCCATGCCAGAAAGGCCGTAACCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACA
CGGTACGGTCTTTCCGGCATTGGGTGTAGTTGAGGCACACCTTTCTGGAAGACCTTCTGT

ValThrProIleAspThrThrIleMetAlaLysAsnGluValPheCysValGlnProGlu
6301 ATGTAACACCAATAGACACTACCATCATGGCTAAGAACGAGGTTTTCTGCGTTCAGCCTG
TACATTGTGGTTATCTGTGATGGTAGTACCGATTCTTGCTCCAAAAGACGCAAGTCGGAC

LysGlyGlyArgLysProAlaArgLeuIleValPheProAspLeuGlyValArgValCys
6361 AGAAGGGGGGTCGTAAGCCAGCTCGTCTCATCGTGTTCCCCGATCTGGGCGTGCGCGTGT
TCTTCCCCCCAGCATTCGGTCGAGCAGAGTAGCACAAGGGGCTAGACCCGCACGCGCACA

GluLysMetAlaLeuTyrAspValValThrLysLeuProLeuAlaValMetGlySerSer
6421 GCGAAAAGATGGCTTTGTACGACGTGGTTACAAAGCTCCCCTTGGCCGTGATGGGAAGCT
CGCTTTTCTACCGAAACATGCTGCACCAATGTTTCGAGGGGAACCGGCACTACCCTTCGA

TyrGlyPheGlnTyrSerProGlyGlnArgValGluPheLeuValGlnAlaTrpLysSer
6481 CCTACGGATTCCAATACTCACCAGGACAGCGGGTTGAATTCCTCGTGCAAGCGTGGAAGT
GGATGCCTAAGGTTATGAGTGGTCCTGTCGCCCAACTTAAGGAGCACGTTCGCACCTTCA

LysLysThrProMetGlyPheSerTyrAspThrArgCysPheAspSerThrValThrGlu
6541 CCAAGAAAACCCCAATGGGGTTCTCGTATGATACCCGCTGCTTTGACTCCACAGTCACTG
GGTTCTTTTGGGGTTACCCCAAGAGCATACTATGGGCGACGAAACTGAGGTGTCAGTGAC

SerAspIleArgThrGluGluAlaIleTyrGlnCysCysAspLeuAspProGlnAlaArg
6601 AGAGCGACATCCGTACGGAGGAGGCAATCTACCAATGTTGTGACCTCGACCCCCAAGCCC
TCTCGCTGTAGGCATGCCTCCTCCGTTAGATGGTTACAACACTGGAGCTGGGGGTTCGGG

ValAlaIleLysSerLeuThrGluArgLeuTyrValGlyGlyProLeuThrAsnSerArg
6661 GCGTGGCCATCAAGTCCCTCACCGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAA
CGCACCGGTAGTTCAGGGAGTGGCTCTCCGAAATACAACCCCCGGGAGAATGGTTAAGTT

GlyGluAsnCysGlyTyrArgArgCysArgAlaSerGlyValLeuThrThrSerCysGly
6721 GGGGGGAGAACTGCGGCTATCGCAGGTGCCGCGCGAGCGGCGTACTGACAACTAGCTGTG
CCCCCCTCTTGACGCCGATAGCGTCCACGGCGCGCTCGCCGCATGACTGTTGATCGACAC

AsnThrLeuThrCysTyrIleLysAlaArgAlaAlaCysArgAlaAlaGlyLeuGlnAsp
6781 GTAACACCCTCACTTGCTACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGG
CATTGTGGGAGTGAACGATGTAGTTCCGGGCCCGTCGGACAGCTCGGCGTCCCGAGGTCC

CysThrMetLeuValCysGlyAspAspLeuValValIleCysGluSerAlaGlyValGln
6841 ACTGCACCATGCTCGTGTGTGGCGACGACTTAGTCGTTATCTGTGAAAGCGCGGGGGTCC
TGACGTGGTACGAGCACACACCGCTGCTGAATCAGCAATAGACACTTTCGCGCCCCCAGG

FIG. 47-7

```
     GluAspAlaAlaSerLeuArgAlaPheThrGluAlaMetThrArgTyrSerAlaProPro
6901 AGGAGGACGCGGCGAGCCTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCC
     TCCTCCTGCGCCGCTCGGACTCTCGGAAGTGCCTCCGATACTGGTCCATGAGGCGGGGGG

     GlyAspProProGlnProGluTyrAspLeuGluLeuIleThrSerCysSerSerAsnVal
6961 CTGGGGACCCCCCACAACCAGAATACGACTTGGAGCTCATAACATCATGCTCCTCCAACG
     GACCCCTGGGGGGTGTTGGTCTTATGCTGAACCTCGAGTATTGTAGTACGAGGAGGTTGC

     SerValAlaHisAspGlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThrThr
7021 TGTCAGTCGCCCACGACGGCGCTGGAAAGAGGGTCTACTACCTCACCCGTGACCCTACAA
     ACAGTCAGCGGGTGCTGCCGCGACCTTTCTCCCAGATGATGGAGTGGGCACTGGGATGTT

     ProLeuAlaArgAlaAlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeuGly
7081 CCCCCCTCGCGAGAGCTGCGTGGGAGACAGCAAGACACACTCCAGTCAATTCCTGGCTAG
     GGGGGGAGCGCTCTCGACGCACCCTCTGTCGTTCTGTGTGAGGTCAGTTAAGGACCGATC

     AsnIleIleMetPheAlaProThrLeuTrpAlaArgMetIleLeuMetThrHisPhePhe
7141 GCAACATAATCATGTTTGCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCT
     CGTTGTATTAGTACAAACGGGGGTGTGACACCCGCTCCTACTATGACTACTGGGTAAAGA

     SerValLeuIleAlaArgAspGlnLeuGluGlnAlaLeuAspCysGluIleTyrGlyAla
7201 TTAGCGTCCTTATAGCCAGGGACCAGCTTGAACAGGCCCTCGATTGCGAGATCTACGGGG
     AATCGCAGGAATATCGGTCCCTGGTCGAACTTGTCCGGGAGCTAACGCTCTAGATGCCCC

     CysTyrSerIleGluProLeuAspLeuProProIleIleGlnArgLeu
7261 CCTGCTACTCCATAGAACCACTTGATCTACCTCCAATCATTCAAAGACTC
     GGACGATGAGGTATCTTGGTGAACTAGATGGAGGTTAGTAAGTTTCTGAG
```

FIG. 47-8

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US88/04125

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

INT. Cl.(4) A61K 39/12; C12Q 1/70; C12N 15/00;C07K 7/10;A61K 37/04; A61K 37/ 02; C07H 15/12.

US Cl. 424/89;435/5;435/6;435/172.2;530/324;530/350;530/387;536/27;

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System	Classification Symbols
U.S.	424/89; 435/5; 435/6; 435/172.2; 530/324; 530/350; 530/387; 536/27;

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [8]

Online search conducted on chemical Abstracts and Biosis.

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category *	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
PX	Journal of Virology, Volume 62 No.2 issued February 1988, Weiner et al "A single antigenomic open frame of the Hepatitis Delta virus encodes the epitope(s) of both Hepatitis Delta antigen polypeptides p 24 Delta and p 27 Delta". See pages 594-599.	1-17,20-21 25,27-28 40
X	Nature, Volume 323, issued 9 October 1986, Wang et al, "Structure, sequence and expression of the Hepatitis Delta viral genome". See pages 508-513	1-17, 20-21 25,27 28,40
A	Journal of Virological Methods, Volum 8 1984, Neurath et al "Strategies for detection of transfusion-transmitted viruses eluding identification by conventional serologic tests. II. Dectection o host DNA in Human Plasmas with elevated alanine aminotransferase" See pages 73-85	22-24 29-31

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search: 03 February 1989

Date of Mailing of this International Search Report: 03 APR 1989

International Searching Authority: ISA/US

Signature of Authorized Officer: Abdel A. Mohamed

Abdel A. Mohamed

International Application No. PCT/US88/04125

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
Y	U.S.A. 4,673,634, Seto et al, Published 16 June, 1987 (Column 1-10)	9-13,20 22-24, 29-34
Y	U.S.A 4,702,909, Villarejos et al, Published 27 October, 1987 (column 1-10)	9-13, 18-20,22-24 26,29,34,39
Y	WO 82/00205, Coursget et al, Published 21 January 1982 "Non-A Non-B Hepatitis Assay and Vaccine" (See the Whole Document)	22-34
Y	WO 87/05930, Foung et al, Published 8 October 1987 "Immortalized Virus-specific tissue cells" (See the whole Document)	35-37

International Application No. PCT/US88/04125

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE [1]

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☐ Claim numbers , because they relate to subject matter [13] not required to be searched by this Authority, namely:

2. ☐ Claim numbers....... , because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out [13], specifically:

3. ☐ Claim numbers____________, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. ☒ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING [2]

This International Searching Authority found multiple inventions in this international application as follows:

See Attached Sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application. Telephone Practice (See Attached Sheet)
2. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:
3. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:
4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

ATTACHMENT TO FORM PCT/ISA/210, Part VI. 1

Telephone Approval:

$280.00 payment is approved by the attorney Gladys H. Monroy on 3 February 1989 for Groups I,II and III; charge to Deposit Account No. 03-1952.

REASONS FOR HOLDING LACK OF UNITY IF INVENTION:

The invention as defined by Group II (claim 32) classified in class 424, subclass 89 and Group III (claim 40) classified in class 435, subclass 172.3, are drawn to a further alternative composition and a further alternative method additional to the compositions and methods of Group I and accordingly, lack of unity of invention under PCT Rule 13.2.

International Application No.: PCT/US88104125

Attachment to Form PCT/ISA/210, Part VI.

VI. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING:

I Claim 1-31 and 33-39, Drawn to purified HCV, polynucleotides thereof, Polypeptide therefore, recombinant vectors, hosts, Antibodies thereto and probes thereto; classified in classes 435,536,530 and 424; subclasses 235,27, 300+ and 89+.

II Claim 32, Drawn to a polypeptide vaccine; classified in class 424, subclass 89.

III Claim 40, drawn to a method of isolating cDNA; classified in class 435, subclass 172.3.